XANTHOMONAS LIVING THE SWEET LIFE: CHARACTERIZATION OF A PLANT SUGAR TRANSPORTER FAMILY AND ITS ROLE IN BACTERIAL BLIGHT OF COTTON SUSCEPTIBILITY

A Dissertation

by

KEVIN LAMARR COX JR.

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Libo Shan
Committee Members,	Michael Kolomiets
	Ping He
	Tom Isakeit
Head of Department,	Leland S. Pierson

December 2018

Major Subject: Plant Pathology

ABSTRACT

Upland cotton (*Gossypium hirsutum)* is an important cash crop that is constantly pressured by various biotic stresses. Bacterial blight of cotton (BBC) caused by *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) is a destructive disease that limits cotton production worldwide. *Xcm* causes disease by injecting transcription activator-like (TAL) effectors in plant cell to directly induce the expression of host genes. Despite having at least 10 TAL effectors cloned from *Xcm*, there have yet to be any susceptibility or resistance proteins identified and characterized. The rapid re-emergence of this disease in the last few years in the US calls for pressing needs to decipher the molecular mechanisms of BBC in order to provide resources to combat this pathogen.

In this dissertation study, I identified a plant sugar transporter (*GhSWEET*) gene family in cotton that plays an important role in BBC susceptibility. To bypass the traditional map-based cloning to identify susceptibility targets, a combination of genome-wide transcriptome profiling and computational prediction of TAL effector binding sites, coupled with functional characterization of candidate genes was used to identify *GhSWEET10* as a key susceptibility target of Avrb6, a major TAL effector in the strain of *Xcm*H1005. This was the first reported susceptibility gene in cotton. Additionally, my data suggest that activating different *GhSWEETs* by a repertoire of TAL effectors is a common, but important, strategy for *Xcm* to cause infection, as current *Xcm* field isolates activated two different *GhSWEETs*, *GhSWEET14a* and *GhSWEET14b*. Further functional analysis on these two *SWEETs* revealed that activating

expression of *GhSWEET14a* and *GhSWEET14b* was correlated with disease susceptibility and genetically required for *Xcm* to cause infection on cotton. Finally, by using genome editing and artificial microRNAs, I was able to develop strategies that could potentially be used to biologically control of BBC. In conclusion, my dissertation revealed the molecular mechanism of BBC and provided a set of toolkits that could be used in the future to develop BBC-tolerant cotton cultivars.

DEDICATION

I dedicate this dissertation to my lovely wife, Charneva Cox, for motivating me to go to graduate school, showing patience, and being my number one supporter, and my daughter, Kyra Cox, for allowing me to interrupt our fun activities at times in order to work in the lab. I also dedicate this dissertation to my father and mother, Kevin Sr. and Gloria Cox, and to my brother, Christopher Cox, for their encouragement and support. Finally, I dedicate this dissertation to my extended family and friends that gave me their support throughout my degree program.

ACKNOWLEDGEMENTS

I would like to sincerely thank my committee chair, Dr. Libo Shan, and her collaborator, Dr. Ping He, for allowing me to conduct my Ph.D. studies in their labs. I'm deeply grateful for their patience, constructive criticisms, and guidance during my time as a graduate student. Their mentoring had a dramatic impact on my development as a scientist. The experiences and opportunities I have encountered would not have been possible without the two of them. I deeply appreciate for all they have done for me and I look forward to applying the mentoring I've received from them to build my career.

I would like to thank my committee members, Dr. Mike Kolomiets and Dr. Tom Isakeit, for their guidance and support throughout the course of this research. I appreciated the interactions and scientific discussions with them, as they provided valuable inputs to significantly improve my research.

I greatly appreciate the collaboration of my work with Dr. Adam Bogdanove and members of his lab. This collaboration has been on-going for nearly my entire tenure as a graduate student. It has been nothing but fruitful and rewarding, as he and his lab conducted key experiments and provided useful inputs to help my projects. I couldn't have some experiments in a timely manner without them.

Many thanks to the current and previous lab members of Dr. Shan's and Dr. He's lab that I've worked with during my graduate studies. They provided great mentoring and scientific discussions during my time in the lab. In addition, they provided fun memories that I will treasure forever. I would like to give a special thanks to Dr.

Fanhong Meng, who collaborated with me on the work in Chapter 2 and provided valuable mentoring that allowed me to significantly improve my lab techniques. Additional thanks to Dr. Fangjun Li, Dr. Ping Wang, and Dr. Xiyu Ma for their help on various experiments presented in this work.

Thanks to the students in the PLPM department for making my graduate studies here enjoyable. I will never forget the times we had together and there are bonds that I have made with some of you that I will cherish forever. I would also like to thank the faculty and staff in the PLPM department, as well as those members of the Institute for Plant Genomics and Biotechnology Center for their helpful interactions. Many thanks to my friends that I made outside the PLPM department for their fun times and memories.

Finally, thanks to my mom, dad, and brother for their continued encouragement and love, my extended family for their support, and to my wife and daughter for their patience and their love. I'm deeply grateful for all of them.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Libo Shan [advisor], Professors Michael Kolomiets and Tom Isakeit of the Department of Plant Pathology & Microbiology, and Professor Ping He of the Department of Biochemistry & Biophysics.

The work in Chapter 2 was performed in collaboration with Dr. Adam Bogdanove's lab of Cornell University and contributed by the authors listed in the publication. The *Xcm*H strains and plasmids were contributed by Dr. Dean Gabriel of the University of Florida. The crop transformation in Chapter 3 was performed by Mr. Marco and Mrs. Mayra Molina of the Laboratory of Crop Transformation located in the Institute of Plant Genomics and Biotechnology Center.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Graduate Diversity Fellowship from Texas A&M University and by a USDA-NIFA AFRI Pre-doctoral Fellowship under Grant Number 2017-67011-26060. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of USDA-NIFA.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW: RETURN OF AN OLD FOE: RE-EMERGENCE OF BACTERIAL BLIGHT OF COTTON

Overview

Biotic stresses can cause a limitation in cotton production every year. Bacterial blight of cotton, a historically controlled disease, has re-emerged in the United States within the last decade for unknown reasons. Since this disease was previously poorly characterized, it is essential to fully understand its interaction with cotton. In this chapter, I will review the current status of the BBC in the field as well as any known molecular mechanisms that have been deciphered. In addition, I will provide insights on how this disease may have re-emerged in cotton fields and possible methods to use current genomic technologies to biologically control of this pathogen.

Introduction

Cotton (*Gossypium* spp.) is an economically important crop and provides the world a significant source of fiber, feed, foodstuff, oil, and biofuel products worldwide. In particular, upland cotton (*Gossypium hirsutum*) provides nearly 95% of those cotton resources. The United States of America is currently the second largest country in the world in upland cotton exports as approximately $7.5 billion dollars were exported by the U.S. in 2017 [1]. However, there are multiple biotic stresses can severely limit cotton production.

One of the major biotic stress of cotton is a destructive bacterial disease called bacterial blight of cotton (BBC). The disease was relatively common several decades ago and caused epidemics to cotton fields. By planting resistant commercial cultivars that were identified through extensive breeding programs and having all seeds undergo a chemical treatment called "acid delinting", the disease has been well controlled since the 1970s [2]. However within the last 8 years, the disease has re-emerged in the U.S. for reasons that have not been clearly revealed. In 2017, BBC caused a yield loss of $20 million dollars [3]. The increase in yield loss in each preceding year has generated a pressing need to understand the molecular mechanisms of this disease in order to employ new genetic material to control these diseases. In this chapter, I will discuss the background of this disease and its current status in cotton fields, the known pathogenicity mechanisms, and provide insights of possible causes of its re-emergence.

Bacterial blight of cotton

BBC, caused by *Xanthomonas citri* subsp. *malvacearum* (*Xcm*), was first identified in the U.S. in 1891. The pathogen can infect all growth stages of cotton. The bacterium is a seedborne pathogen, but can also enter through the stomates of leaves or open wounds from infected crop debris from a previous season [4]. The early symptom of this disease starts off as water-soaked lesions on leaves and then progress into angular lesions due to the leaf veins restricting movement. The disease becomes more severe as it progresses, as the leaf veins blacken causing a "blighting" appearance, lesions appear on cotton bolls and begin to rotten, and pre-mature defoliation occurs. BBC can quickly

spread to other areas of the field through wind-driven rain or irrigation. There currently no effective chemical treatments that can be applied once the disease is established in the field.

Xcm is classified into at least 22 races (named from race 1 to race 22) according to its virulence on a selection of cotton cultivars [2]. If a strain overcomes a certain resistance gene in a cotton cultivar and causes susceptibility, that strain is considered to be another race. These races tend to be geographically distinct, as they can be found in India and countries in Africa. The predominant race in the U.S. is race 18, which was first identified in Texas in 1973 [5, 6].

In order to cause disease, *Xcm* injects effector proteins into plant cells via the type III secretion system to suppress plant immunity and promote pathogenicity [7]. One distinct class of effectors in *Xcm* are transcription activator-like (TAL) effectors, a class highly conserved among *Xanthomonas* spp., that functionally resemble eukaryotic transcription factors [8]. These effectors consists of central tandem repeats that can bind to the promoters of host genes to drive transcription of genes involved in susceptibility (known as *S* genes) or resistance (known as *R* genes) [9]. The decoding of host DNA recognition by TAL effectors has made it possible to computationally predict the TAL effector's binding site in a host genome and identify candidate target genes [10] (see chapter II for more details).

A common class of *S* genes that different *Xanthomonas* species target in different pathosystems are *SWEET* genes, a plant sugar transporter family. In rice, *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) can use several TAL effectors to induce *OsSWEET11*,

OsSWEET13, and *OsSWEET14* to cause bacterial leaf blight [11-14]. In cassava, *Xanthomonas axonopodis* pv. *manihotis* uses $TAL20_{Xam668}$ to induce *MeSWEET10a* to cause bacterial blight of cassava [15]. A similar mechanism was revealed in cotton, as *Xcm* uses a TAL effector, Avrb6, to drive expression of a cotton *SWEET* gene, *GhSWEET10*, to cause BBC susceptibility [16]. This suggests that *Xcm* redirects sugar in cotton to facilitate its nutrition (Figure 1). To our knowledge, this is the only reported TAL effector from *Xcm* that has its target gene identified.

While TAL effectors can directly activate *S* or *R* genes, there is also a class of TAL effectors that can suppress resistance, known as truncated TAL effectors (truncTALEs) [17]. Interestingly, these truncTALEs do not have DNA binding activity as a typical TAL effector, but instead suppress the activity of disease resistance proteins, likely through a direct protein-protein interaction. This has been observed in rice, as the hypersensitive response triggered by the *Xa1* resistance locus is suppressed by two *Xoo* truncTALEs, Tal3a and Tal3b [18], and the HR triggered by the *Xo1* resistance locus is suppressed by an *Xanthomonas oryzae* pv. *oryzicola* truncTALE, Tal2h [17]. It is possible that this recently discovered class of TAL effectors exist in *Xcm*. The TAL effector family in *Xcm* field isolates needs to be studied more extensively as they appear to be key virulence factors in causing disease.

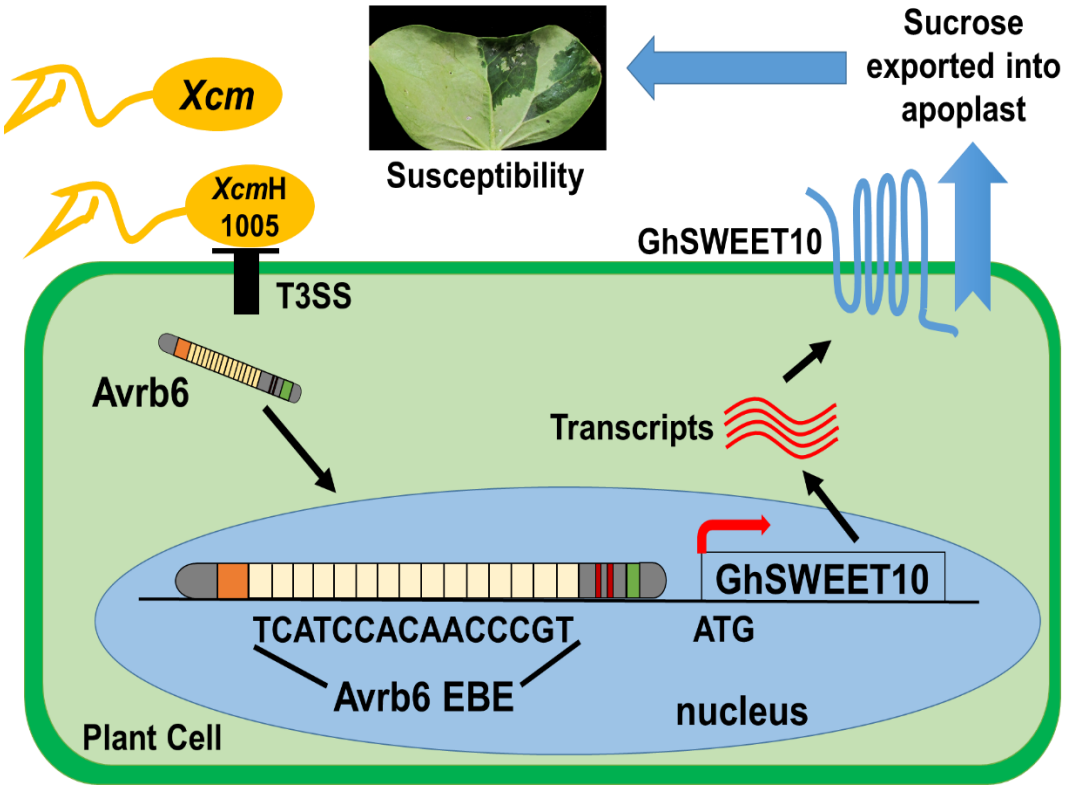

Figure 1. A model on how *Xcm*H1005 causes BBC susceptibility.
Upon entry into the plant, *Xcm*H1005 injects TAL effector Avrb6 into plant cell via its type III secretion system (T3SS). Avrb6 localizes to nucleus where it binds to the promoter of *GhSWEET10*, specifically the Avrb6 effector binding element (EBE), to activate transcription of *GhSWEET10*. The abundance of *GhSWEET10* transcripts could cause an over-accumulation of sucrose being exported into apoplast. This could allow *Xcm* to use this excess sucrose as a carbon source for fitness and survival to mediate BBC susceptibility.

In addition to TAL effectors, *Xanthomonas* species can also use other type III effectors, or non-TALEs (also known as *Xanthomonas* outer proteins, Xops), to suppress a host defense response and cause susceptibility [19, 20]. Qin et al. demonstrated in a recent publication that *Xanthomonas* outer protein K (XopK), an E3 ubiquitin-ligase, from *Xoo* is required for virulence as it inhibits PTI upstream of mitogen-activated protein kinase cascades in rice [21]. Specifically, XopK interacts and directly ubiquitinates rice OsSERK2, a key signaling component in immunity and development, to cause degradation via the proteasome in order to disrupt its activity. There has not been any Xops functionally characterized in *Xcm*. Therefore, non-TALEs in *Xcm* should be investigated to determine if there are any effectors involved in suppressing immunity. These non-TALEs could be used to suppress R proteins in cotton in order to prevent its TAL effector from being recognized while targeting a host gene.

While the mechanisms of BBC are now starting to be unraveled, the cause of the re-emergence of this disease in the U.S. has not been clearly deciphered. BBC was thoroughly controlled since the 1970s until an outbreak occurred in 2011 in Arkansas, Mississippi, and Missouri and then in other cotton growing states in the preceding years. One possible explanation of this re-emergence as described by Phillips et al. is the increase of susceptible cultivars being planted in fields [22]. These particular cultivars may have favorable traits for farmers such as resistance to nematodes, another destructive cotton pathogen, and better fiber quality. This could have allowed a low population of *Xcm* to survive for years until it was able to cause an epidemic including on resistance cultivars. Another possible reason for the re-emergence is that there could

have been polymorphisms in the TAL effectors of *Xcm*. This could allow the TAL effectors to upregulate a new *S* gene to break resistance. Additionally, the polymorphisms could also prevent the TAL effectors from activating an *R* gene that would lead to a defense response. Interestingly, isolates that were gathered in Texas during the 2015 season strongly activated two new *SWEET* members, *GhSWEET14a* and *GhSWEET14b*, instead of *GhSWEET10*, suggesting a change in the TAL effector assemblies between the *Xcm* strain harboring Avrb6 and the current field isolates [16]. In order to further study and identify TAL effectors from these *Xcm* field isolates, additional long read DNA sequencing platform, such as Pacific Biosciences (PacBio) and MinION sequencing, needs to be employed. To date, there are only two studies that have used long read sequencing platforms on *Xcm* strains [16, 23]. Revealing the TAL effector assemblies on these relatively newer isolates will provide a blueprint to perform pathogenesis studies on the *Xcm*-cotton pathosystem and reveal the cause of this re-emergence.

Targeting veins to hijack the water source of plants

With BBC re-emerging, there has been an increased desire to fully understand how *Xcm* infect cotton. One key characteristic of BBC is the symptoms of it can be identified around the veins of the leaves, which are responsible for transporting water throughout the plant. During the later stages of BBC infection, the veins around the infected leaves begin to blacken [24]. A possible reasoning for this is *Xcm* could potentially use the veins as a primary water source to accelerate the infection process.

This could potentially lead to the veins leaking water and presenting a “blackening” phenotype.

The water-soaking symptom of BBC can be identified in any part of the infected leaf but interestingly, it appears to be more centered on the veins. The veins serves as a “highway” for water and minerals to be transported through the leaf and onto the rest of the plant. During the later stages of BBC infection, the veins on the infected leaf begins to blacken and appear as “black lightning bolts”. This observation suggests that having an aqueous habitat is a key component during *Xcm* infection, similar to *Pseudomonas syringae* during its infection in *Arabidopsis* [25]. But whether creating this aqueous environment also promotes bacterial multiplication, survival, or dispersal is still unclear. In addition, the direct cause of water-soaking has not been fully resolved. Since upregulating *GhSWEET* genes is an important component for *Xcm* for causing infection, these plant sugar transporters could play a role in the water-soaking symptom. By inducing *GhSWEET* genes during infection, this would not only redirect sugars to the bacteria to feed on as a carbon source, but also change the osmotic potential in the apoplast [26]. This could cause excess water to accumulate in the apoplast area, resulting it to leak onto the exterior surface of the infected leaf. Investigating the mechanisms of water-soaking caused by *Xcm* will provide hints on how the pathogen establishes and sustains itself in the early stages of infection.

Increase in humidity and temperature plays a role in disease severity

The disease triangle model remains as a core principal of plant pathology as it shows the interactions between the environment, the pathogen, and the plant. In order for disease to occur, there must be an optimal environment, a virulent pathogen, and a susceptible plant. Therefore, a virulent pathogen theoretically cannot infect a susceptible host if the environment is not favorable. In contrast, an optimal environment can enhance disease. With global temperatures increasing due to climate change, pathogens now have more favorable conditions to cause epidemics. Notably, BBC has historically been known to be more destructive to cotton fields in climates with higher temperatures and humidity [4, 24].

The stomates are natural openings on leaves that allows for gas exchange. This also provides bacteria an entry site to penetrate the leaf and enter the apoplast. To defend themselves, plants have evolved signaling cascades that are activated upon the detection of bacteria to force stomatal closure [27]. This disrupts pathogen entry and prevents the disease from being severe or fatal to the plant [28]. However, this stomatal closure can be overridden when the humidity is high [29], which has been observed more often in recent years. Since this high humidity climate causes the stomates to remain open, this potentially provides more opportunities for *Xcm* to enter the plant more easily and subsequently cause infection.

The increase in temperature and humidity globally has also been promoting heavier rainfall within the last several years. The rainfalls are more common and severe during the warmer seasons as warmer air contains more moisture available to storms.

The consequence of this are cotton fields that are damp and humid, a very optimal environment for BBC to develop and spread rapidly. The changing environmental conditions over the last several years could have created the optimal conditions for re-emergence of this pathogen. Climate change appears to be having an impact on crop diseases and needs to be studied extensively with these cotton pathogens to investigate their effects on crop yield.

Molecular strategies to breed resistant cultivars

The rise of a powerful genome editing platform called clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein 9 (Cas9) has provided the ability to perform functional analysis of genes through loss-of-function or gain-of-function of gene expression [30]. Recent studies have shown that transforming CRISPR/Cas9 constructs into cotton is feasible [31-34] with improved editing efficiency [35]. A strategy to combat against BBC would be to edit key susceptibility genes in order to prevent them being activated by the pathogen, specifically with *GhSWEET* genes as they appear to be key susceptibility targets. In rice, genome editing has been used to modify *OsSWEET* susceptibility genes to prevent them from being activated by *Xoo* during infection [11, 36]. Since *SWEET* genes are important to the development of the plant, new approaches of CRISPR are being applied in which the TAL effector binding sites found in the promoter region of targeted *SWEET* genes are being edited in order to prevent activation by the TAL effector without disrupting the endogenous function of the host's sugar transporter.

Plant disease resistance (*R*) genes provide a robust counter attack against pathogens [37]. However, *R* genes have not been well-characterized in cotton, especially for BBC. This likely due to the complexity of the tetraploid genome of upland cotton. Despite these shortcomings, there have been several quantitative trait loci (QTLs) identified that appear to contribute to resistance. In some examples, the cotton line Acb6 harbors a QTL that carries a recessive resistance locus called "b6" that triggers an HR response upon the detection of the Avrb6 effector [38]. This appears to be a classic "gene-for-gene" interaction, as the *avrb6* mutant is able to cause water-soaking on Acb6 without triggering HR. There are also multiple cotton lines that have "*B*" genes that provided resistance against specific avirulence genes of *Xcm*. One of these genes, B_{12}, confers resistance to all races of *Xcm* in the U.S., including the commonly found race 18 [39]. Another cotton line, Im216, is also reported to be resistant to all races of *Xcm* [2, 40]. However, none of the genes localized on these QTLs have been identified or functionally characterized. The recent releases of draft genomes sequences for upland cotton will enable future studies to identify the responsible genes involved in resistance in these cultivars and reveal the mechanisms on how they trigger resistance. This would allow breeders and researchers to produce more durable cotton resistance against this pathogen.

Conclusions

The re-emergence of BBC came as a surprise to several cotton growers in the U.S. This facilitated the need to understand the biology of the *Xcm*-cotton pathosystem.

While our knowledge on this is still relatively unclear, the recent findings on this disease have begun to shed light on the mechanisms of BBC susceptibility and the reasoning for its resurgence. In addition, the recent advancements in genetic material in cotton research provides promise that this disease could be biologically controlled once we understand its progression. Continuing to study this re-emerging pathogen will be critical in order to increase our knowledge and provide cotton growers cultivars with increased resistance to BBC.

CHAPTER II

TAL EFFECTOR DRIVEN INDUCTION OF A *SWEET* GENE CONFERS SUSCEPTIBILITY TO BACTERIAL BLIGHT OF COTTON[*]

Overview

Transcription activator-like (TAL) effectors from *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) are essential for bacterial blight of cotton (BBC). Here, by combining transcriptome profiling with TAL effector binding element (EBE) prediction, we show that *GhSWEET10*, encoding a functional sucrose transporter, is induced by Avrb6, a TAL effector determining *Xcm* pathogenicity. Activation of *GhSWEET10* by designer TAL effectors restores virulence of *Xcm avrb6* deletion strains whereas silencing of *GhSWEET10* compromises cotton susceptibility to infections. A BBC-resistant line carrying an unknown recessive *b6* gene bears the same EBE as the susceptible line, but Avrb6-mediated induction of *GhSWEET10* is reduced, suggesting a unique mechanism underlying *b6*-mediated resistance. We show via an extensive survey of *GhSWEET* transcriptional responsiveness to different *Xcm* field isolates that additional *GhSWEETs* may also be involved in BBC. These findings advance our understanding of the disease and resistance in cotton and may facilitate the development cotton with improved resistance to BBC.

[*] Reprinted from [121].

Introduction

Cotton (*Gossypium* spp.) is an economically important crop and provides a significant source of fiber, feed, foodstuff, oil, and biofuel products worldwide. The cotton genus is composed of at least 45 diploid and 5 tetraploid species[41]. The tetraploid (AD genome) species, including *G. hirsutum* that produces 95% of the world's cotton fiber, is likely derived from a hybridization between an A-genome-like ancestral species resembling *G. arboreum* and a D-genome-like ancestral species resembling *G. raimondii*[42]. The availability of the draft genome sequences for *G. raimondii, G. arboreum, G. hirsutum* and *G. barbadense* not only provides genetic resources to study the complex genome evolution and polyploidization process but also lays the foundation for functional genomic approaches to dissect cotton gene functions with a goal to improve its agricultural performance in the face of biotic and abiotic stresses[43-48].

Bacterial blight of cotton (BBC), caused by *Xanthomonas citri* subsp. *malvacearum* (*Xcm*), is among the destructive diseases of cotton[2]. Following epidemics in the 1970s, the disease has occurred sporadically in the US, but in the past several years re-emerged as a significant yield constraint. There are pressing needs to address the underlying mechanisms of susceptibility and resistance to BBC in cotton. *Xcm* injects effector proteins into plant cells via the type III secretion system to promote pathogenicity in plants. In the presence of corresponding resistance (R) proteins, some of these effectors trigger resistance and function as avirulence proteins. Interestingly, all known pathogenicity and avirulence factors (encoded by so called "*pth*" and "*avr*" genes) of *Xcm* are transcription activator-like (TAL) effectors[49]. TAL effectors

functionally resemble eukaryotic transcription factors and upregulate host genes by directly binding to their promoters[50, 51]. TAL effectors are highly conserved among different *Xanthomonas* spp., with an N-terminal type III translocation signal, a central repeat region (CRR), and C-terminal nuclear localization signals (NLS) followed by an acidic activation domain (AD). The proteins differ mainly in the CRR, which consists of 1.5-33.5 copies of near perfect repeats of 33-34 amino acids. These repeats are conserved with the exception of the 12th and 13th residues of each copy, defined as the repeat variable di-residue (RVD)[52, 53]. Each RVD targets a specific nucleotide of the promoters of host genes, creating a code, such that the sequence of RVDs defines the effector binding element (EBE); each EBE starts, however, with a nearly invariant thymine, which is specified in a yet unclear manner by structures immediately N-terminus of the CRR[52-55].

Among the identified TAL effector targets include a group of pepper genes up-regulated by AvrBs3 from *X. campestris* pv. *vesicatoria* (*Xcv*)[56, 57]. These include the *R* gene *Bs3*, and *UPA20*, which encodes a basic helix-loop-helix (bHLH) transcription factor regulating plant cell hypertrophy[58]. *Os8N3*/*Xa13/OsSWEET11*, essential for reproductive development, is a rice susceptibility (*S*) gene targeted by TAL effector PthXo1 from *X. oryzae* pv. *oryzae* (*Xoo*)[14]. The *xa13* allele, which is unable to be induced by PthXo1, acts as a recessive *r* gene against *Xoo* infection[59, 60]. Notably, four *Os8N3*/*Xa13/OsSWEET11* homologs, rice *Os11N3/OsSWEET14* and *Xa25*/*OsSWEET13*, pepper *UPA16*, and cassava *MeSWEET10a* are also targeted by TAL effectors from *Xoo*, *Xcv*, and *X. axonopodis* pv. *manihotis* (*Xam*) respectively[13, 15, 61]. It was hypothesized

that *Xanthomonas spp.* induces the expression of *SWEET* genes during infection to transport sucrose to the apoplast, thereby providing the bacteria with a carbon source[11-15, 62-64]. TAL effectors from *X. citri sp citri* (*Xcc*), the causal agent for citrus bacterial canker, induce the expression of *CsSWEET1* in citrus, but the *CsSWEET1* does not contribute demonstrably to susceptibility; another TAL effector target however, *CsLOB1*, does function as an *S* gene, promoting the characteristic pustule formation and bacterial multiplication[65]. Other characterized TAL effector targets include the rice bacterial blight *R* genes *Xa27* [66], *Xa23* ([67], and *Xa10* [68], the bacterial leaf streak of rice *S* gene *OsSULTR3;6* ([69] and the pepper bacterial spot *R* gene *Bs4C* [70].

Previous work identified at least 10 TAL effectors from the cotton pathogen *Xcm*H1005 and showed that a mutant with a deletion of 7 *tal* genes was no longer able to cause observable water-soaking symptoms on Acala44 (Ac44), a BBC susceptible line of cotton. Avrb6 from *Xcm*H1005 was among the first examples of an individual TAL effector important for *Xcm* virulence. An *avrb6* mutant of *Xcm*H1005, *Xcm*H1407 (*Xcm*H1005Δ*avrb6*), showed reduced water-soaking whereas when expressed in a relatively weak virulent strain *Xcm*N1003, Avrb6 enhances water-soaking[71]. Although it does not affect *in planta* bacterial multiplication, Avrb6 plays a major role in release of bacteria from the leaf interior to the leaf surface during infections[71].

To identify the TAL effectors responsible for virulence and their targets, we sequenced the whole genomes of *Xcm* strains *Xcm*H1005 and *Xcm*N1003 and assembled the full repertoire of TAL effectors. Further, we deployed genome-wide gene expression profiling in cotton coupled with TAL effector-DNA binding code-assisted EBE

prediction and identified the clade III sucrose transporter gene *GhSWEET10* as a target of Avrb6. We determined that *GhSWEET10* is a major *S* gene for BBC by using designer TAL effectors (dTALEs) to induce it independently of any other possible Avrb6 targets and by silencing it with *Agrobacterium*-mediated virus-induced gene silencing (VIGS). Significantly, other members of the clade III *GhSWEET* genes are strongly induced by different *Xcm* field strains responsible for recent re-emergence of the disease in the southern United States. Our data indicate that cotton GhSWEET10 and likely other SWEET sugar transporters in the same clade play a major role in BBC and suggest that newly evolved or horizontally transferred TAL effectors targeting these genes attribute to the re-emergence of BBC in the field. In addition, a BBC-resistant cotton line carrying *b6*, a genetically complex resistance gene that has not yet been molecularly identified, showed markedly reduced Avrb6-mediated induction of *GhSWEET10*, ostensibly due to promoter polymorphisms outside the EBE. This suggests a novel mechanism for resistance that may represent a genetic determinant of *b6.*

Results

Genome and TAL effector sequences of Xcm *H1005 and N1003*

The strains *Xcm*H1005, a rifamycin resistant derivative of *Xcm*H, and *Xcm*N1003, a spectinomycin and rifamycin resistant derivative of *Xcm*N, are two *Xcm* strains that cause water-soaking on Ac44 cotton[38]. To set the stage for identifying the role that TAL effectors in these strains play in BBC, we generated complete genome sequences using Single Molecule Real Time sequencing (Pacific Biosciences; hereafter

'PacBio sequencing'). The H1005 genome consists of a circular chromosome of 5,212,564 bp and a plasmid, pXcmH [72] of 88,283 bp. The N1003 genome consists of a circular chromosome of 5,218,607 bp and a plasmid, herein designated as pXcmN, of 59,644 bp. The general features of each genome are presented in Table 1. The overall architectures of the two chromosomes are highly similar, with a single, large inversion (Figure 2a). The plasmids are dissimilar, with only a few, relatively small regions of homology (Figure 2a). The *tal* genes of the two strains are largely distinct (Figure 2b and Supplementary Table 2). H1005 encodes 12 TAL effectors with six in the chromosome and six in pXcmH. Ten of these are Avr proteins named previously based on reactions of differential cotton lines to transformants of an otherwise compatible *Xcm* strain carrying corresponding cosmid subclones: AvrB4, AvrB5 Avrb6, Avrb7, AvrB101, AvrB102, AvrB103, AvrB104, AvrBIn, and AvrBn [38, 72, 73]. Another is PthN2, originally characterized in *Xcm*N1003 [49]. The twelfth is uncharacterized, and we designated it as $Tal6_{XcmH1005}$, following a previously described naming scheme[74]. *Xcm*N1003 harbors nine *tal* genes, five on the chromosome and four on pXcmN. Four on the chromosome and one on the plasmid however are disrupted by a frameshift mutation or a large insertion in the repeat region or the 3' end of the gene. One of the disrupted genes, *tal1'*, on the chromosome, carries only a last repeat and is otherwise complete until an intergrase insertion further downstream. Another of the disrupted genes, *pthN*' on pXcmN, is otherwise identical to *pthN*, an *Xcm*N1003 *tal* gene shown previously to contribute to water-soaking without triggering resistance on commercial U.S. cotton varieties[49]. The intact *pthN* gene is also located on pXcmN, along with *pthN2*, which

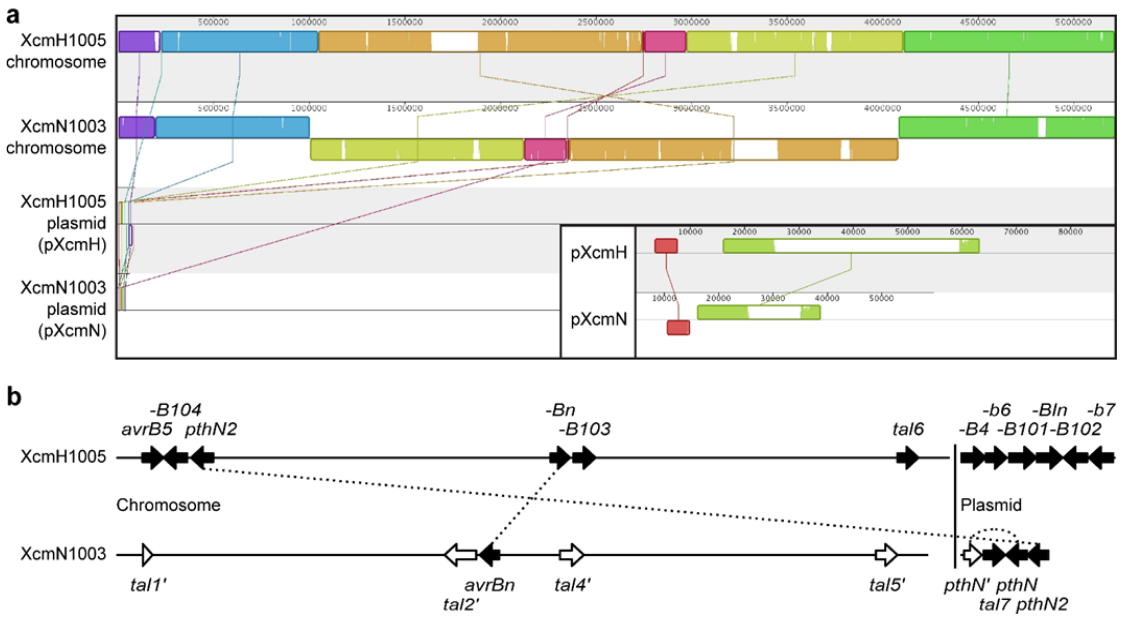

Figure 2. Comparison of whole genomes and *tal* genes of *Xcm*H1005 and *Xcm*N1003.

(A) Alignment of the *Xcm*H1005 and *Xcm*N1003 genomes, generated using progressiveMAUVE with default parameters[75]. Each genome comprises a single circular chromosome and a single circular plasmid, shown linearized. Colored, rounded rectangles represent locally collinear blocks (LCB), regions of homology without rearrangement across the aligned sequences, connected by matching colored diagonal lines. The orientations of the LCB, forward or reverse, are indicated by their position above or below the line, respectively. The height of a column within a block reflects the average similarity relative to the other aligned sequence(s) there (see http://darlinglab.org/mauve/user-guide/viewer.html for details). Inset (bottom right) shows the alignment of the plasmids only, at a larger scale. Horizontal axes show sequence coordinates (bp). MAUVE backbone files giving the exact coordinates of all LCB are provided as Supplementary Table 1 (plasmids only). (B) The *tal* genes of *Xcm*H1005 and *Xcm*N1003. The genes are represented as block arrows at their relative positions in the chromosome or plasmid (horizontal lines). The arrows are magnified relative to the rest of the genome, but intergenic regions and arrow sizes relative to each other are to scale. Dashed lines connecting two arrows indicate identical encoded RVD sequences. Gene names follow the scheme of [74], except for the indicated *avr* and *pth* genes, named previously[38, 49, 72, 73]. For label clarity, hypens replace "*avr*". An apostrophe following the gene name (white arrows) indicates that the coding sequence is terminated early due to a frameshift mutation or other coding sequence disruption. Details, including RVD sequences and AnnoTALE designations [76] are given in Supplementary Table 2.

Table 1. Genome characteristics of *Xcm*H1005 and *Xcm*N1003

Characteristics	*Xcm* H1005	*Xcm* N1003
Length (bp)	5,212,564 bp	5,218,607 bp
No. of protein-coding genes	4,227	4,218
GC content	64.62%	64.54%
No. of tRNA genes	54	54
No. of rRNA operons	2	2
Components	circular chromosome (5.2 Mb); plasmid (88,283 bp)	circular chromosome (5.2 Mb); plasmid (59,644 bp)
No. of TAL effectors (Chromosome::Plasmid)	6::6	5::4
GenBank Accession No.	CP013004	CP013006

also contributes to water-soaking, but less strongly[49]. Another intact *tal* gene, *tal7*$_{XcmN1003}$, also resides on the plasmid. The fourth intact gene is *avrBn* located on the chromosome. Apart from *pthN'*, *pthN*, *pthN2*, and *avrBn*, no other *Xcm*N1003 *tal* genes show obvious similarity in their RVD sequences to any other *tal* genes in either strain. The divergence in *tal* gene content between the two strains, both with respect to RVD composition and genomic location of orthologs, is striking, which is in line with the fact that they were isolated from geographically distant regions[38].

Candidate targets of Avrb6

Avrb6 is a TAL effector in *Xcm*H1005 that causes strong water-soaking in the cotton line Ac44 [71]. In this study, we used the Ac44E tetraploid genotype, which is a single plant selection from Brinkerhoff's original Ac44 parent[77]. Ac44E has similar morphology to Ac44 except it has more abundant flowers and is slightly more susceptible to BBC than Ac44 [77]. To identify the cotton genes that are specifically activated by Avrb6, we used an integrative approach involving a whole-genome RNA-sequencing (RNA-Seq) analysis of protoplasts of Ac44E expressing Avrb6 and computational prediction of Avrb6 EBEs in cotton (Figure 3a). Compared to the whole plant-pathogen infection assay, the homogenous protoplast mesophyll cell system expressing the individual bacterial effector circumvents the complication of multiple pathogen effectors and elicitors that could spontaneously activate or suppress a large number of host genes. We cloned the full-length coding sequence of Avrb6 from *Xcm*H1005 into a plant expression vector with an HA-epitope tag at the C-terminus

under the control of the *35S* CaMV promoter. The sequence was confirmed via Sanger sequencing (Figure A-1). An immunoblot analysis with an α-HA antibody detected a major polypeptide slightly above the 100kDa marker, which matches the predicted molecular mass of 108 kDa for Avrb6 (Figure 3b). We further sub-cloned the *35S::avrb6-HA* expression cassette into a binary vector and performed *Agrobacterium*-mediated transient expression assays in Ac44E cotton and *Nicotiana benthamiana*. As shown in Figure 3c, we observed a similar result as in cotton protoplasts with a major polypeptide being detected at ~108 kDa. Next, we performed RNA-Seq analysis with cotton protoplasts 12 hrs after transfection with the *avrb6* construct or a vector control. Approximately 39-50 million raw read pairs and 36-46 million cleaned read pairs were obtained for each sample, which corresponds to ~450 x coverage of 77,267 annotated protein-coding transcripts[41]. Using a two-fold expression change and an adjusted *p*-value < 0.05 as cut-offs, and using the *G. raimondii* genome sequence as a reference (the tetraploid genome sequence was not available at the time), we identified 2,026 genes that were induced by Avrb6 (Figure 3d). To identify genes with potential EBEs for Avrb6, we used the TALE-NT 2.0 Target Finder tool to examine all gene promoter sequences in the *G. raimondii* genome, using the standard score ratio cutoff of 3.0 (Figure 3e). The list of genes displaying one or more putative EBEs was then intersected with the list of genes up-regulated by Avrb6 to produce a list of candidate of Avrb6 targets (Table 2). These candidates were ranked by the probability that the predicted EBE is functional by using a machine learning algorithm based on the output of TALE-NT 2.0 and the genomic context of the predicted EBEs[69]. Interestingly, among the top candidates,

Gorai.008G209000.1, a homolog of the SWEET sucrose transporter genes that serve as *S* genes in rice and cassava, was induced ~6000-fold by Avrb6 and has an EBE with a probability of 1 (Table 2). Other top candidates include Gorai.007G067700.1, Gorai.010G056300.1, Gorai.009G327300.1, and Gorai.008G047400.1. We then used the gene IDs from *G. raimondii* to identify their homologs in *G. hirsutum*. The gene IDs of the homologs identified in *G. hirsutum* are Gh_D12G1898 for Gorai.008G209000.1, Gh_D11G0631 for Gorai.007G067700.1, Gh_D06G0459 for Gorai.010G056300.1, Gh_D05G2954 for Gorai.009G327300.1, and Gh_D12G0420 for Gorai.008G047400.1. These genes were named *GhSWEET10D* (see below for explanation), *Gh067700*, *GhKBS1*, *GhHLH1*, and *GhMDR1* respectively based on their predicted gene annotation.

Direct induction of targets by Avrb6

To independently determine the induction of the candidate genes, we performed reverse transcription (RT)-PCR assays using Ac44E cotton protoplasts transfected with Avrb6 or with an empty vector as control. The top five candidates, *GhSWEET10*, *Gh067700*, *GhKBS1*, *GhHLH1*, and *GhMDR1*, were strongly induced by Avrb6 compared to the vector control (Figure 4a). We further confirmed induction in cotton leaves by inoculating *Xcm* strains with or without *avrb6*. HM2.2S is derived from H1005 with a deletion of at least seven *tal* genes including *avrb6*, while HM2.2S (*avrb6*) is HM2.2S carrying *avrb6* in plasmid vector pUFR127 [49]. H1005Δ*avrb6* is an *avrb6* mutant derivative of *Xcm*H1005 [71]. As previously reported, *avrb6* in different *Xcm*

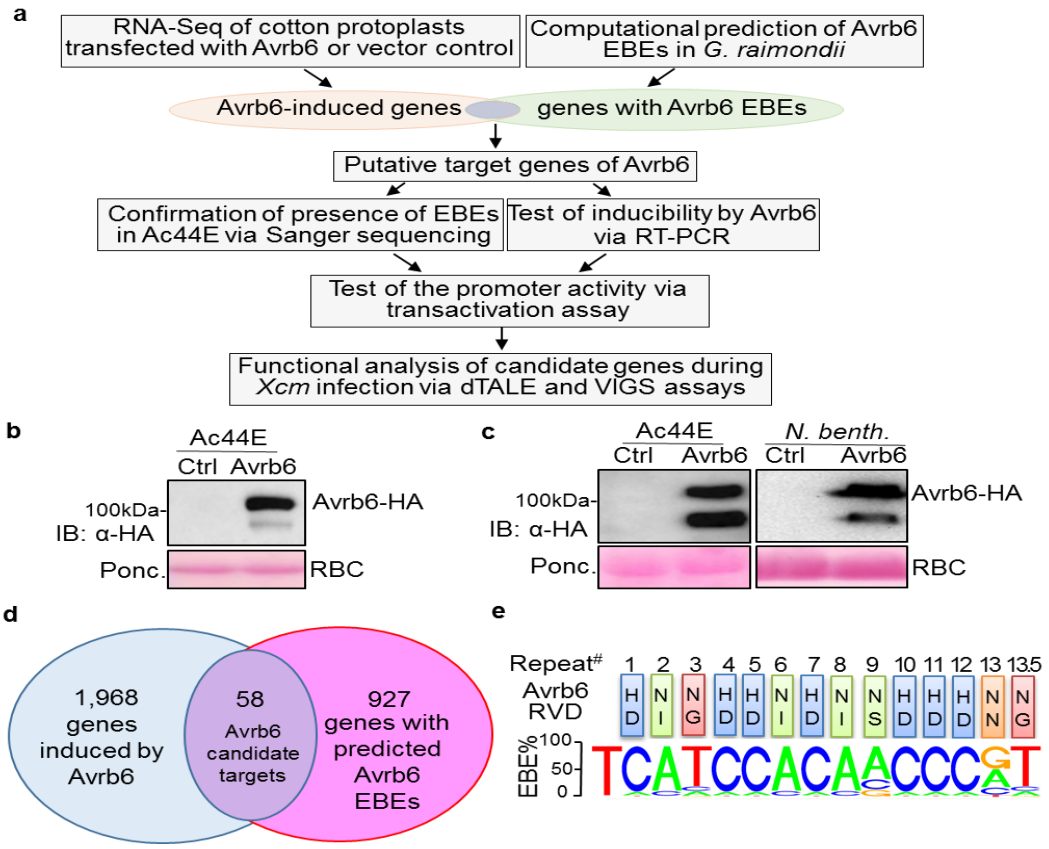

Figure 3. Transcriptome profiling coupled with EBE prediction reveals candidate target genes of Avrb6.

(a) Schematic diagram of experimental design to identify Avrb6 target genes in cotton. (b) Avrb6 protein expression in cotton protoplasts. Cotton protoplasts isolated from Ac44E were transfected with *avrb6-HA* or an empty vector as a control (Ctrl). Samples were harvested 12 h after transfection and subjected to immunoblotting with α–HA antibody (top panel). Ponceau S. staining (Ponc.) of total protein served as the protein loading control; RuBisCo (RBC) is shown (bottom panel). (c) Avrb6 protein expression in cotton and *N. benthamiana*. Cotyledons from two-week-old Ac44E cotton and leaves of four-week-old *N. benthamiana* were infiltrated with *Agrobacterium* carrying *35S::avrb6-HA* or an empty vector control (Ctrl). Inoculated tissues were harvested at 48 hpi and subjected to immunoblotting. B and C were repeated three times with similar results. (d) Venn diagram of Avrb6-induced genes and genes with Avrb6 EBEs in *G. raimondii* genome. (e) The predicted EBE of Avrb6 RVDs. Colored boxes on the top panel display the RVD of each repeat of Avrb6. Bottom panel indicates the relative frequencies of RVD associations with the four nucleotides for each repeat.

Table 2. The list of top 10 candidate target genes of Avrb6 in cotton based on their inducibility by Avrb6 and the probability of Avrb6 EBEs

Gene ID	Gene Name	Induction fold	EBE Prob.	Annotation
Gorai.008G209000.1	*GhSWEET10D*	5973.1	1.0	Bidirectional sugar transporter N3
Gorai.007G067700.1	*Gh067700*	29.2	0.97	Unknown protein
Gorai.010G056300.1	*GhKBS1*	9.6	0.96	Kinase family protein
Gorai.009G327300.1	*GhHLH1*	4.0	0.89	Basic helix-loop helix TF
Gorai.008G047400.1	*GhMDR1*	7.5	0.71	Multi-drug resistance ABC transporter
Gorai.010G246000.1	*Gh2460*	5.6	0.65	Unknown protein
Gorai.007G152300.1	*GhPPR1*	5.2	0.64	Pentatricopeptide repeat-containing protein
Gorai.008G222500.1	*GhCYP1*	471.6	0.59	Cytochrome P450
Gorai.001G131000.2	*GhZFP1*	12.3	0.51	Zinc finger family protein
Gorai.009G045100.2	*GhGLY1*	12.8	0.5	Glycine dehydrogenase

strains induced strong water-soaking on Ac44E cotton[71] (Figure 4b). We observed that water-soaking was only induced by the *Xcm* strains carrying *avrb6* [HM2.2S (*avrb6*), HM2.2S (*placZ*::*avrb6*) - a strain that contains the plasmid pUFR135 which expresses *avrb6* driven by the *lacZ* promoter, H1005 (*placZ*::*avrb6*), and H1005]. As shown in Figure 4c, the induction of *GhSWEET10*, *Gh067700*, *GhKBS1*, and *GhMDR1* by the strains carrying *avrb6*, HM2.2S (*avrb6*) and H1005, was stronger than that by strains lacking *avrb6*, HM2.2S and H1005Δ*avrb6*, at 12- and 24-hr post inoculation (hpi). We could not amplify the *GhHLH1* gene in cotton likely due to low transcript level. To explore whether *GhSWEET10*, *Gh067700*, *GhKBS1*, and *GhMDR1* are directly induced by Avrb6, we added the eukaryotic protein synthesis inhibitor cycloheximide (CHX; 50 μM final concentration) to the *Xcm* suspensions and inoculated into Ac44E cotton leaves. Each gene was still induced by strains carrying *avrb6*, H1005 and HM2.2S (*avrb6*), in the presence of CHX at 24 hpi (Figure 4d), suggesting that new protein biosynthesis is not required for induction of these genes and that they are direct targets of Avrb6. The effectiveness of CHX in suppressing protein synthesis in this context was confirmed by its suppression of Avrb6 protein expression in a separate *Agrobacterium*-mediated transient expression assay in Ac44E cotton (Figure A-2). To examine potential polymorphisms in these gene promoters between the sequenced cotton and Ac44E cotton used in this study, we amplified each promoter from Ac44E for sequencing. This revealed that each of the promoters possesses an Avrb6 EBE except *Gh067700*, which has a 1 bp deletion in the EBE in Ac44E (Figure A-3). To confirm whether Avrb6 activates the genes by targeting the EBEs, we amplified ~800 bp upstream of the

translational start site of each gene, cloned each of these fragments upstream of a luciferase reporter gene (*LUC*) and performed a cotton protoplast-based transactivation assay (Figure 3e). Compared to an empty vector control, co-transfection of Avrb6 strongly activated *pGhSWEET10D::LUC* with about 110-fold induction, and also activated *pGhKBS1*::*LUC* and *pGhMDR1::LUC* with about 15~20-fold induction (Figure 3f). Significantly, these promoters were not activated by PthN, another *Xcm* TAL effector with an RVD sequence distinct from Avrb6, indicating specificity of activation by Avrb6 (Figure 3f). Consistent with the observation that *Gh067700* in cotton Ac44E carries a nucleotide deletion in Avrb6 EBE, *pGh067700::LUC* was not activated by Avrb6. In addition, *pGhHLH1::LUC* was not significantly activated by Avrb6, though its promoter possesses a predicted Avrb6 EBE. Since the promoter of *GhSWEET10D* was strongly induced by Avrb6, we tested the role of the EBE by mutating two nucleotides to create a version of *pGhSWEET10D::LUC* with a "CA" to "GG" substitution in the EBE (*mEBE*; Figure 3g). Avrb6 strongly activated *pGhSWEET10D::LUC* carrying the wild-type EBE, but not the *mEBE* (Figure 3h). Taken together, these data indicate that Avrb6 directly and specifically activates *GhSWEET10D*, and likely *GhKBS1* and *GhMDR1*, and that the EBE is required for Avrb6-induced expression of *GhSWEET10D*.

GhSWEET10 *activation using a dTALE induces water-soaking*

We generated dTALEs to specifically activate *GhSWEET10D*, *GhKBS1*, or *GhMDR1* via EBEs distinct from the corresponding Avrb6 EBEs, introduced these into

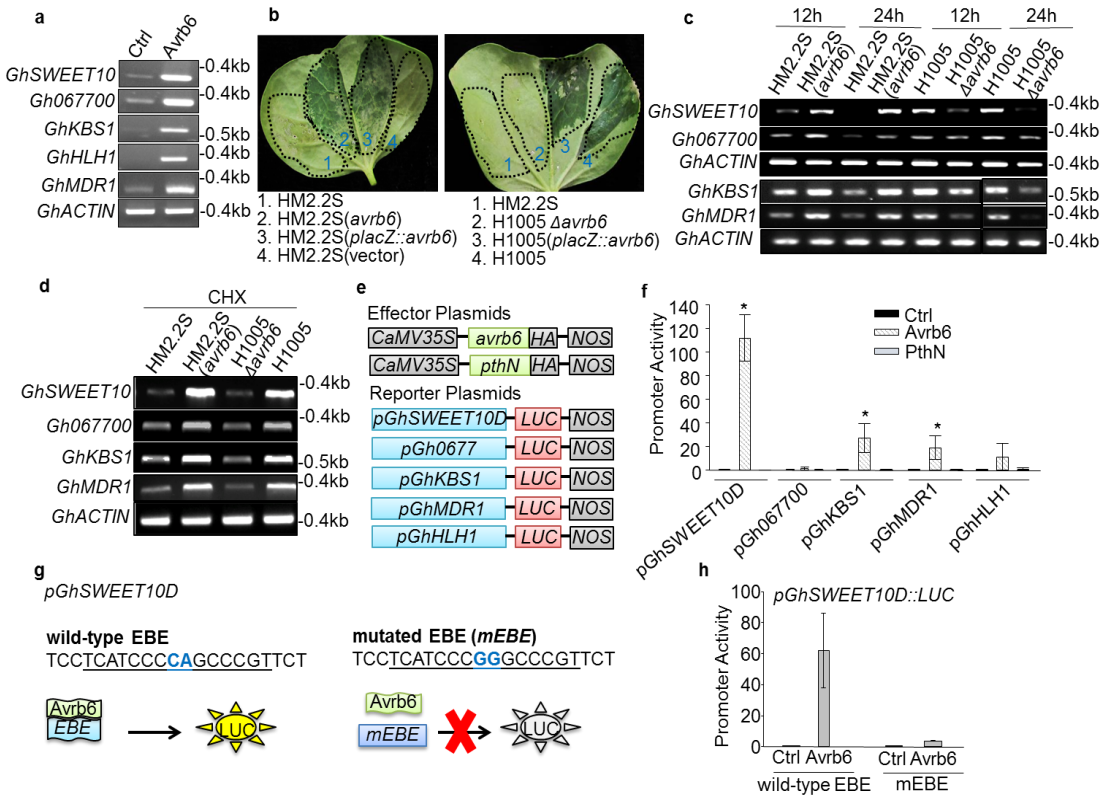

Figure 4. Avrb6 directly induces cotton gene transcription in the absence of protein synthesis.
(a) RT-PCR analysis of Avrb6 up-regulated genes in cotton protoplasts. Cotton protoplasts of Ac44E were transfected with *avrb6-HA* or a vector control (Ctrl) and incubated for 12 h before RNA isolation. *GhACTIN* was used as an internal control. (b) Avrb6 contributes to water-soaking development in cotton. Cotyledons from two-week-old Ac44E cotton were syringe-inoculated with different *Xcm* strains at OD_{600}=0.1 and photographed four days after infiltration. Inoculation areas are indicated by dotted lines. (c) RT-PCR analysis of Avrb6 up-regulated genes in cotton upon *Xcm* infection. Cotyledons from two-week-old Ac44E cotton were syringe-inoculated with different *Xcm* strains. Tissues were harvested at 12 and 24 hpi. (d) RT-PCR analysis of Avrb6 up-regulated genes in the presence of cycloheximide (CHX). Cotyledons from two-week-old cotton were syringe-inoculated with different *Xcm* strains in 50 μM CHX. Tissues were harvested at 24 hpi for RT-PCR. (e) Schematic diagram of the effector and reporter constructs. The reporter construct contains an expression cassette with a luciferase (*LUC*) reporter gene under the control of a candidate gene promoter. The effector construct contains either Avrb6 or PthN with an HA epitope tag under the control of the CaMV *35S* promoter. (f) Transcriptional activity of Avrb6 in cotton protoplasts. Protoplasts were co-transfected with a reporter construct and *avrb6*, *pthN*, or a vector control (Ctrl), and were collected 12 h after transfection. *UBQ10-GUS* was included in the transfections as an internal control. The luciferase activity was normalized with GUS activity. The data are shown as mean ± SD (n=3) from three independent repeats. Asterisks indicate significant difference using two-tailed t-test ($p<0.05$). (g) Schematic diagram of the transactivation assay of wild-type and mutated EBE (mEBE) in *pGhSWEET10D* in response to Avrb6. The nucleotide sequence containing the putative EBE is shown and the two nucleotides that were mutated are highlighted in blue. (h) Transcriptional activity of *GhSWEET10D* with wild-type and mEBE in response to Avrb6. Cotton protoplasts were co-transfected with *pGhSWEET10D::LUC* carrying wild-type or mEBE and Avrb6 or a vector control (Ctrl). The data are shown as mean ± SD (n=3) from three independent repeats. The above experiments were repeated three times with similar results.

the *Xcm* HM2.2S strain, which lacks *avrb6*, and performed disease assays in cotton. The dTALEs were expressed from the low-copy, broad-host range vector, pKEB1, a derivative we constructed from pUFR047 [38] that carries a Gateway cloning cassette (Figure 5a). Significantly, HM2.2S transformants carrying the dTALE targeting *GhSWEET10D* (dTALE1), but not those targeting *GhMDR1* (dTALE3) or *GhKBS1* (dTALE4), or the vector control, induced water-soaking in Ac44E cotton, similarly to *Xcm*H1005 (Figure 5b). To determine if the dTALEs used indeed activate the transcripts of their respective target genes, we performed RT-PCR analysis. As shown in Fig. 5c, *GhSWEET10D* and *GhMDR1* were induced by the cognate dTALEs. These data indicate that activation of *GhSWEET10D* upon *Xcm* infection contributes to water-soaking development, thus identifying *GhSWEET10D* as a relevant target of Avrb6 in cotton and an *S* gene. We did not detect the induction of *GhKBS1* by its dTALE, and cannot conclude whether GhKBS1 plays a role in BBC.

Avrb6 targets GhSWEET10 *in both A and D genomes in cotton*

The tetraploid *G. hirsutum* has two subgenomes, A and D. The above studies of *GhSWEET10D* were mainly based on the D genome of *G. raimondii* and the corresponding D subgenome of *G. hirsutum.* We examined if Avrb6 also targets *SWEET10* in the A-genome of *G. arboreum,* which is closest to the progenitor species of the A-subgenome of *G. hirsutum.* We identified a homolog of GhSWEET10D, named GaSWEET10, which bears 98% similarity to GhSWEET10D at the amino acid level. We expressed Avrb6 in *G. arboreum* cotton protoplasts and determined whether

GaSWEET10 was induced by Avrb6. RT-PCR analysis showed that *GaSWEET10* was induced by Avrb6 (Figure 5d). To determine whether Avrb6 could specifically activate the promoter of *GaSWEET10*, we cloned the promoter, *pGaSWEET10*, upstream of the luciferase gene for a transactivation assay. As shown in Fig. 5e, Avrb6 activated *pGaSWEET10::LUC* to a similar level as Avrb6 activation of *pGhSWEET10D::LUC.* Notably, an alignment of the Avrb6 EBE from *pGhSWEET10D* with the one from *pGaSWEET10* revealed them to be identical except for a single nucleotide polymorphism near the 3' end of the EBE (position 14, counting the initial thymine), where mismatches are known to be generally tolerated[78]. Additionally, an alignment of protein coding sequences revealed that GrSWEET10 from *G. raimondii* is identical to D-subgenome GhSWEET10D, whereas GaSWEET10 from *G. arboreum* is identical to A-subgenome GhSWEET10A (Figure A-4). Similarly, the promoter sequence, including the EBE, of *pGhSWEET10A* from tetraploid Ac44E was more closely related to *pGaSWEET10* from *G. arboreum* versus *pGrSWEET10* from *G. raimondii* (Figure A-4b). We used the dTALE approach to determine if *GaSWEET10* is also a direct target of Avrb6 that causes water-soaking. As shown in Figure 5f, the dTALE2 construct, but not the dTALE1 construct specific to *pGhSWEET10D* or an empty vector introduced in H1005Δ*avrb6* induced water-soaking in *G. arboreum*. Notably, H1005 induced relatively weak water-soaking on *G. arboreum* compared to *G. hirsutum*, likely due to *G. arboreum* carrying some strain-specific resistance as noted previously[79]. Taken together, these results indicate that Avrb6 can directly target

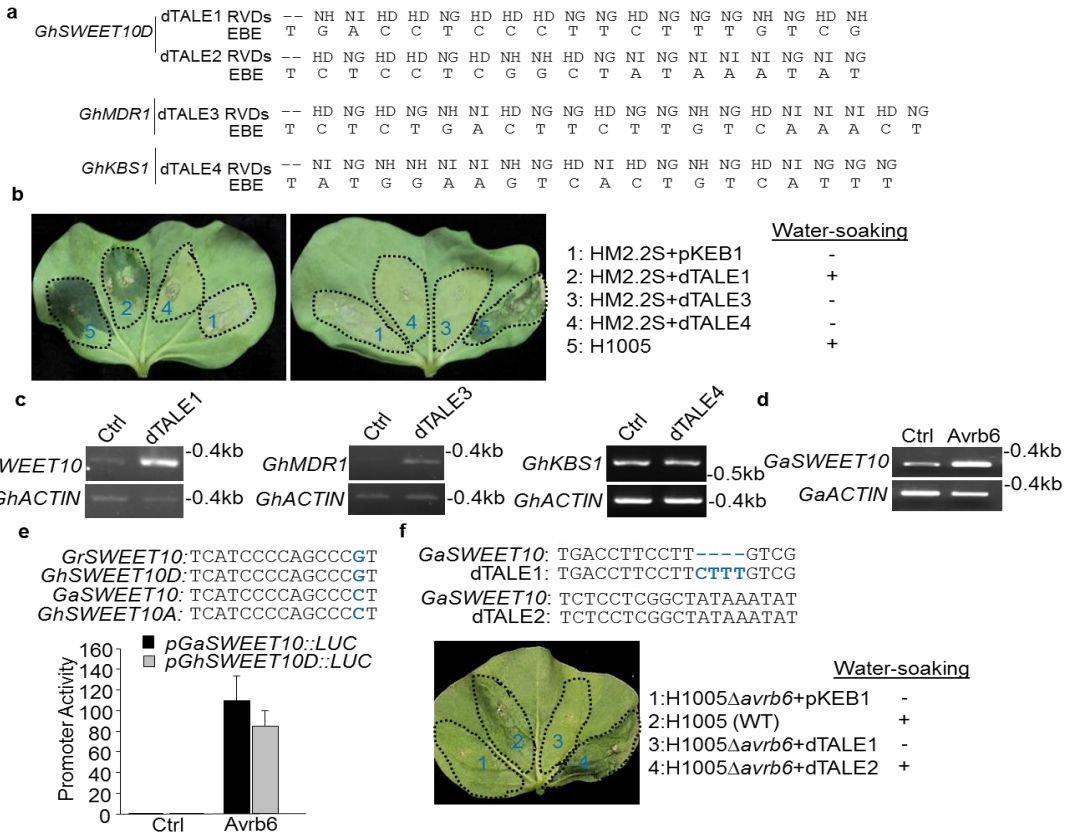

Figure 5. The dTALE activating *GhSWEET10* induces water-soaking in cotton.
(a) RVDs of dTALEs and the corresponding targeted EBE sequences in the *G. hirsutum* genome. (b) *Xcm*HM2.2S expressing a dTALE uniquely targeting *GhSWEET10D* induces water-soaking in cotton. Cotyledons from two-week-old Ac44E plants were syringe-inoculated with different *Xcm* strains at OD_{600}=0.1 and photographed at 4 dpi. The table displays the presence (+) or absence (-) of water-soaking. (c) RT-PCR analysis of *GhSWEET10*, *GhKBS1*, and *GhMDR1* targeted by their respective dTALEs. Two-week-old cotyledons of Ac44E were syringe-inoculated with *Xcm* strains HM2.2S (OD_{600}=0.1) containing different dTALEs. Inoculated tissues were harvested at 24 hpi for RT-PCR analysis. (d) *GaSWEET10* is induced by Avrb6 in *G. arboreum* protoplasts. Protoplasts isolated from *G. arboreum* were transfected with Avrb6 or an empty vector control (Ctrl) and samples were harvested 12 h after transfection for RT-PCR analysis. (e) Transactivation assay of *GaSWEET10* and *GhSWEET10D* promoters in response to Avrb6 in cotton protoplasts. Alignment of Avrb6 EBEs reveals a polymorphism between *GhSWEET10D* and *GaSWEET10* promoters which was highlighted in blue. Protoplasts isolated from commercial cotton variety FM706V were co-transfected with *pGaSWEET10::LUC* or *pGhSWEET10D::LUC* and Avrb6 or a vector control (Ctrl). The data are shown as mean ± SD (n=3) from three independent repeats. (f) The dTALE matching the *GaSWEET10* promoter causes water-soaking on *G. arboreum*. The EBE sequences of two dTALEs targeted to different regions of the *GaSWEET10* promoter are shown. dTALE2 with an EBE sequence perfectly matching the *GaSWEET10* promoter but not dTALE1 with a partially matching EBE sequence restores *Xcm*-mediated water-soaking on cotton. Cotyledons from two-week-old *G. arboreum* were syringe-inoculated with different *Xcm* strains at OD_{600}=0.1 and photographed at 4 dpi. The above experiments were repeated three times with similar results.

SWEET10 in both the A and D genomes to contribute to the development of BBC.

GhSWEET10 *silencing reduces water-soaking caused by* Xcm

Having determined that *GhSWEET10* mediates Avrb6-induced water-soaking, we subsequently investigated whether *GhSWEET10* is genetically required for water-soaking during *Xcm* infection in cotton. We employed the *Agrobacterium*-mediated virus-induced gene silencing (VIGS) system to silence *GhSWEET10* and examined BBC development after subsequent inoculation of *Xcm*. The VIGS construct was designed to silence both *GhSWEET10A* and *GhSWEET10D* because of sequence conservation. There was no apparent difference in plant growth between the VIGS-*GhSWEET10* plants and plants inoculated with the GFP control (VIGS-Ctrl) up to three weeks after VIGS (Figure A-5). Three weeks after *Agrobacterium*-infiltration, the silenced plants were vacuum-infiltrated with *Xcm*H1005. Compared to the control, the VIGS-*GhSWEET10* plants consistently showed fewer water-soaked lesions upon *Xcm*H1005 infection (Figure 6a, 6b). RT-PCR analysis confirmed reduced induction of *GhSWEET10* in VIGS-*GhSWEET10* plants compared to VIGS-Ctrl plants (Fig. 6c). Taken together, the data indicate that silencing of *GhSWEET10* in cotton reduces the development of water-soaking following *Xcm* inoculation, providing genetic evidence that *GhSWEET10* is functionally required for normal BBC symptom development upon infection by *Xcm* strains carrying *avrb6*.

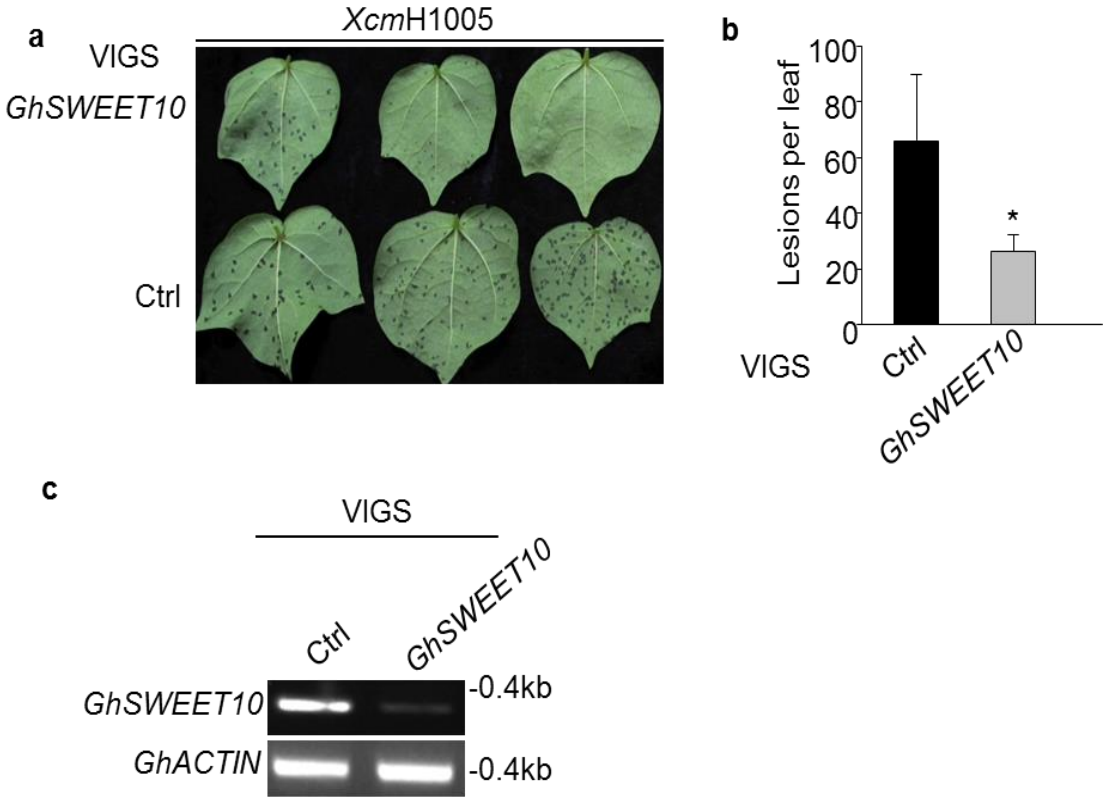

Figure 6. Silencing of *GhSWEET10* in cotton reduces water-soaking caused by *Xcm* infection.

(a) *Xcm*-mediated water-soaking lesion development on cotton plants upon VIGS. Cotyledons from two-week-old Ac44E plants were syringe-infiltrated with *Agrobacterium* carrying a VIGS construct to silence *GhSWEET10* (VIGS-*GhSWEET10*) or a GFP vector control (VIGS-Ctrl). Three weeks later, plants were vacuum-infiltrated with *Xcm*H1005 at OD_{600}=0.01. Images were taken two weeks after inoculation. (b) Quantitative analysis of *Xcm*-mediated water-soaking lesions on cotton plants upon VIGS. Lesions were counted on the 2nd true leaf of each VIGS plant (at least 10 plants were inoculated for each construct in each repeat). The data are shown as mean ± SE (n=10). The asterisk indicates a significant difference with a Student's *t* test ($p < 0.05$) when compared with the control. (c) RT-PCR analysis of *GhSWEET10* expression in cotton plants upon VIGS. VIGS assays were done similarly as in (a) and three weeks later, the 2nd true leaf was harvested for RT-PCR analysis before *Xcm* inoculation. *GhACTIN* was used as an internal control. The above experiments were repeated three times with similar results.

GhSWEET10D *encodes a functional sucrose transporter*

Since *GhSWEET10* belongs to the *SWEET* gene family, we examined whether GhSWEET10 functions as a sugar transporter. To measure activity, we transiently expressed the coding sequence of GhSWEET10D along with a Förster Resonance Energy Transfer (FRET) sucrose sensor in human embryonic kidney T293 (HEK293T) cells[80]. Compared to the vector control, expression of GhSWEET10D enabled HEK293T cells to accumulate sucrose, which was detected as a negative ratio change, to a level comparable to the well-characterized *Arabidopsis* transporter AtSWEET11 (Figure 7a). In addition, when we transiently expressed *GhSWEET10D* under the control of the *35S* CaMV promoter in *N. benthamiana*, the sucrose concentration of the apoplastic fluids from the *35S::GhSWEET10D* inoculated leaves was higher than that from the vector control inoculated leaves (Figure A-6a), indicating that GhSWEET10D is able to transport sucrose *in planta*. Interestingly, *GhSWEET10D* did not restore the growth defect of yeast EBY4000, a hexose transport-deficient yeast strain, on the medium supplemented with 2% glucose or 2% fructose as the sole carbon source, indicating that GhSWEET10D does not mediate efficient fructose or glucose transport (Figure A-6b). Taken together, the data suggested that *GhSWEET10D* encodes a functional sucrose transporter.

GhSWEET10D *is a member of a large gene family in cotton*

In bacterial blight of rice, several closely related *SWEET* genes can each serve as an *S* gene[12]. To better understand the relationship within the SWEET family in cotton,

especially which SWEETs are most closely related to GhSWEET10, the full-length protein sequence of GhSWEET10D was used as the query for a BLASTp analysis against the *G. hirsutum* NBI protein database (www.cottongen.org/tools/blast). The search identified 50 additional GhSWEETs, for a total of 51 SWEETs in *G. hirsutum*, with each possessing the conserved domains found in the AtSWEET superfamily in *Arabidopsis* (Figures A-7, A-8). Based on the well-characterized phylogenetic analysis of AtSWEETs, we generated a phylogenetic tree and classified the GhSWEET family into four clades with nine in clade I, 14 in clade II, 18 in clade III, and 10 in clade IV. The individual members of GhSWEETs are named using the tree described by Eom et al, 2015 as a reference where GhSWEET1-3 are in clade I, GhSWEET4-8 are in clade II, GhSWEET9-15 are in clade III, and GhSWEET16-18 are in clade IV[81]. Considering the gene duplication and polyploidy in tetraploid cotton, lowercase and uppercase letters are used respectively to distinguish closely related members. If multiple members are co-orthologous to the closest AtSWEET, lowercase letters (a-c) are added after the numbers to differentiate genes in the same subclade. An uppercase "A" is further added to indicate the gene is from the A subgenome, or an uppercase "D" for the D subgenome. For example, there appear to be two pairs of GhSWEETs that are closely related to AtSWEET1 as shown in the phylogenetic tree. Therefore, we distinguish these by naming one pair as GhSWEET1a and another pair as GhSWEET1b. Each of these pairs is further named according to their subgenome locations as GhSWEET1aA and GhSWEET1aD, and GhSWEET1bA and GhSWEET1bD respectively (Figure 7b, Figure A-7). Like the *SWEET* genes in rice and cassava that serve as *S* genes,

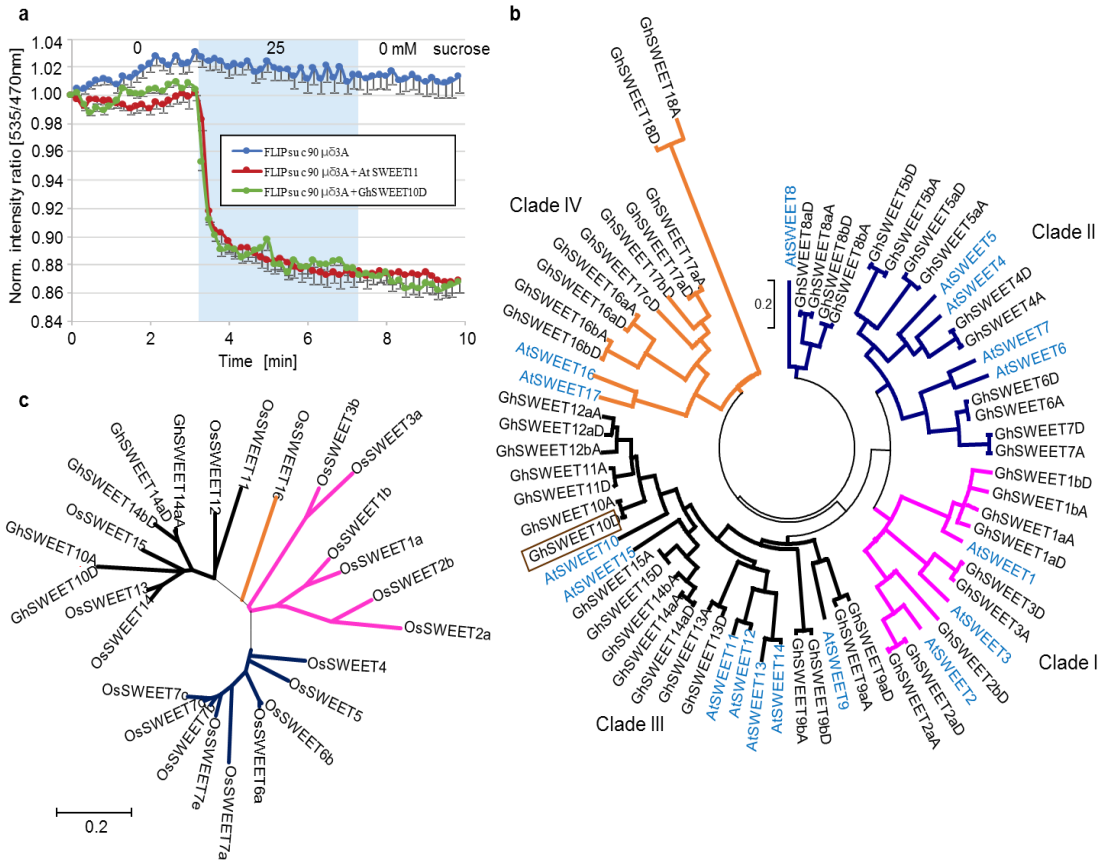

Figure 7. *GhSWEET10D* encodes a functional clade III sucrose transporter.
(a) Detection of GhSWEET10D transporter activity in HEK293T cells using the sucrose sensor FLIPsuc90µδ3A. Sucrose transporter activity was assayed by co-expressing GhSWEET10D with the cytosolic FRET sucrose sensor FLIPsuc90µδ3A in HEK293T cells. Blue circles correspond to cells expressing the sensor alone; red circles and green circles correspond to cells co-expressing the sensor with AtSWEET11 or GhSWEET10D respectively. The cyan block indicates duration of perfusion with 25 mM sucrose. Accumulation of sucrose is reported as a negative ratio change (mean-s.e.m.; n>10). Experiments were repeated four times with similar results. (b) Phylogenetic analysis of GhSWEET proteins from *G. hirsutum*. AtSWEET proteins from *Arabidopsis thaliana* were included to represent different clades of SWEET superfamily. The phylogenetic tree was made using the neighbor-joining method in MEGA version 6.06. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site[82]. *GhSWEET10D*, which was used as the query, is highlighted with a brown box. (c) Phylogenetic analysis of GhSWEET10, GhSWEET14a, GhSWEET14b, and the OsSWEET proteins from *Oryza sativa*. The phylogenetic tree was made as described in b. The clades of the SWEET family were color-coded as follows: clade I=pink, clade II=blue, clade III=black, clade IV=orange.

GhSWEET10 is a member of the clade III family (Fig. 6c).

GhSWEET10 *induction by Avrb6 is reduced in* b6 *cotton*

Avrb6 causes severe water-soaking in the susceptible Ac44E cotton line. However, in the near-isogenic resistant line Acb6, which harbors the apparently genetically complex BBC recessive resistance locus *b6*, Avrb6 causes the resistance-associated hypersensitive response (HR) instead ([38] and Figures A-9a, A-9b). We examined whether *b6*-mediated resistance is in part due to the inability of Avrb6 to activate *GhSWEET10* in Acb6. When we expressed Avrb6 in Ac44E and Acb6 protoplasts respectively, the induction of *GhSWEET10* by Avrb6 was significantly lower in Acb6 than that in Ac44E (Figure 8a). We further compared *GhSWEET10* induction in Ac44E and Acb6 cotton plants inoculated with different *Xcm* strains with or without *avrb6*. Both Ac44E and Acb6 plants showed increased induction of *GhSWEET10* upon inoculation with HM2.2S carrying *avrb6* compared to the plants inoculated with HM2.2S (Figure 8b). However, the induction was significantly reduced in Acb6 plants relative to Ac44E plants. A similar trend was observed when H1005 was used to inoculate both lines (Figure 8b). To examine whether the reduced induction by Avrb6 in Acb6 could be due to polymorphisms at the *GhSWEET10D* EBE, we sequenced the *GhSWEET10D* promoter region from Acb6. The EBE is perfectly conserved, but in Acb6, there is a SNP immediately adjacent to the 5' of the EBE and a 2 bp insertion surrounded by six additional SNPs approximately 20 bp downstream of the EBE (Figure A-9c). We cloned the promoter of *GhSWEET10D* from Acb6 in front of the luciferase

reporter for a transactivation assay in cotton protoplasts. Similar to what we observed for the endogenous *GhSWEET10D* gene, the induction of the *pGhSWEET10D*$_{Acb6}$*::LUC* reporter by Avrb6 was nearly three folds lower than that of the reporter carrying the promoter from Ac44E (*pGhSWEET10D*$_{Ac44E}$*::LUC*) (Figure 8c). The promoter of *GhSWEET10A* is almost identical in Ac44E and Acb6 (Figure A-4b). Thus, the polymorphisms outside the *GhSWEET10D* EBE may impair full induction by Avrb6. In addition, when we inoculated H1005Δ*avrb6 carrying* dTALE2, which activates *pGhSWEET10D* (Figures 4a, 4f), into cotyledons of Acb6, the bacterium triggered water-soaking, but not HR in Acb6 (Figure A-9d). The data suggest that *GhSWEET10D* may not mediate Avrb6-induced HR and Avrb6 targets another gene for HR in Acb6.

Clade III GhSWEET *responsiveness to different* Xcm *isolates*

Clade III *SWEET* genes have been implicated in disease susceptibility in rice and cassava. We here identified a clade III *SWEET* gene in cotton, *GhSWEET10*, that encodes a sucrose transporter and that also plays a role in disease susceptibility. To examine how general the involvement of clade III *GhSWEET* genes in BBC is, we investigated the transcriptional changes of these genes in response to a nine *Xcm* isolates collected from infected cotton fields from different locations in Texas. Notably, these isolates are suspected to be the causal agents of the recent rampant reemergence of BBC in the U.S.. However, it has not been determined yet whether these isolates represent a new race. We found that the field isolates were unable to induce *GhSWEET10*. However,

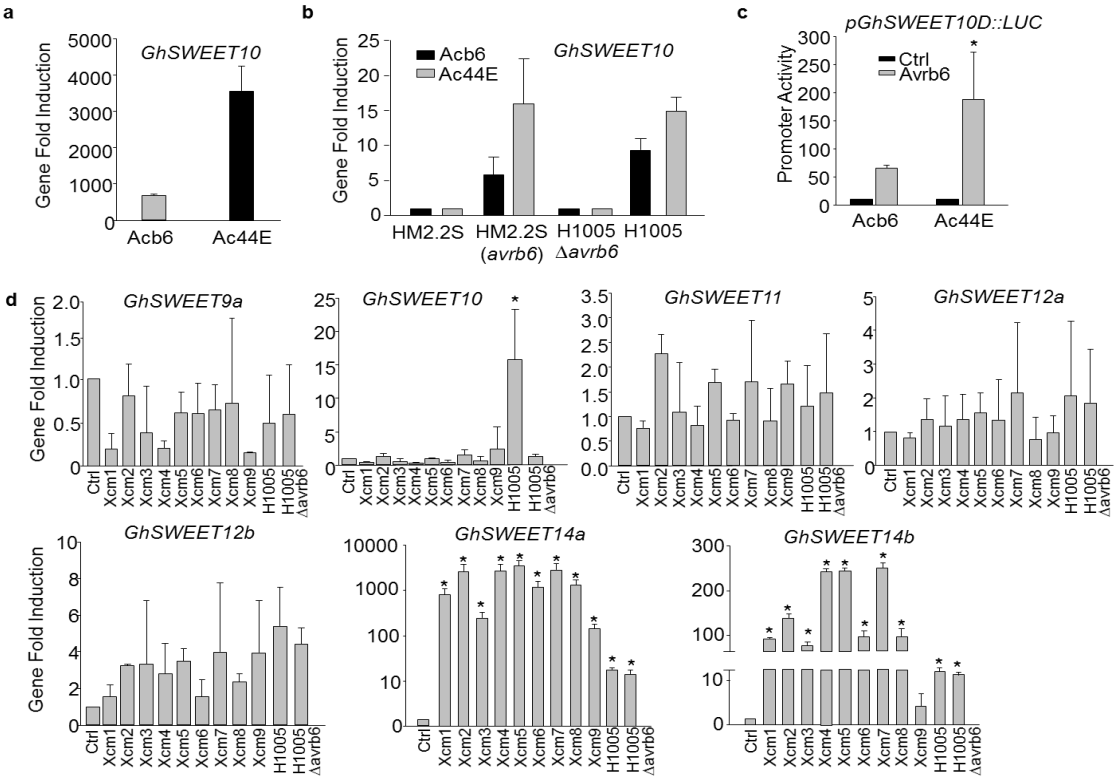

Figure 8. Induction of *GhSWEET* genes in different cotton-*Xcm* interactions.
(a) *GhSWEET10* induction by Avrb6 in protoplasts of the resistant line Acb6 and the susceptible line Ac44E. Graph displays Avrb6-induced *GhSWEET10* reads per kilobase of transcript per million mapped reads (RPKM) and fold change from RNA-Seq analysis with Ac44E (compatible) and Acb6 (incompatible) cotton lines. (b) qRT-PCR analysis of *GhSWEET10* in Ac44E and Acb6 plants upon *Xcm* infections. Two-week-old cotyledons were syringe-inoculated with different *Xcm* strains at OD600=0.1 and tissues were collected at 24 hpi. *GhUBQ1* was used as an internal control. The data are shown as mean ± SD (n=3) from three independent repeats. (c) Transactivation assay of *pGhSWEET10D::LUC* from Acb6 and Ac44E in response to Avrb6 in cotton protoplasts. The promoter of *GhSWEET10D* was amplified from Acb6 and Ac44E respectively and fused with luciferase reporter. Protoplasts of cotton variety FM706V were co-transfected with the reporter construct and Avrb6 or an empty vector control (Ctrl). The data are shown as mean ± SD (n=3) from three independent repeats. An asterisk indicates the significant difference using two-tailed t-test ($p<0.05$) between two reporter constructs after Avrb6 induction. (d) Induction of clade III *GhSWEET* genes by different *Xcm* isolates. Two-week-old cotyledons were syringe-inoculated with different *Xcm* isolates (*Xcm1-Xcm9*) collected from different locations in Texas, U.S. at OD600=0.1. Tissues were collected at 24 hpi. *GhUBQ1* was used as an internal control for qRT-PCR analysis. The data are shown as mean ± SD (n=3) from three independent repeats. An asterisk indicates significant difference using two-tailed t-test ($p<0.01$) compared to the water control. The above experiments were repeated three times with comparable results.

all of the field isolates highly induced *GhSWEET14a* (200-4000 fold) (Fig. 8d). In addition, all of the isolates, except for *Xcm9*, also significantly induced *GhSWEET14b* (20-300 fold). As expected, H1005, but not H1005Δ*avrb6*, induced *GhSWEET10* by 15-fold. Notably, *GhSWEET14a* and *GhSWEET14b* are clade III *SWEET* genes (Fig. 7c). Taken together, these data suggest that in addition to *GhSWEET10*, *GhSWEET14a* and *GhSWEET14b* are potential *S* genes, and that activating clade III *GhSWEET* genes might be broadly important to the ability of different *Xcm* strains to cause BBC.

Discussion

BBC caused by *Xcm* is among devastating diseases that limit cotton yields. Despite several TAL effectors being cloned from *Xcm*, no *R* or *S* genes have yet been molecularly identified, mainly because of the large and polyploidy nature of the cotton genome and limited molecular and genomic tools. The approach taken in this study successfully overcame these challenges to establish a role for SWEET-mediated sugar transporter in BBC. *GhSWEET10* was strongly induced by Avrb6 during *Xcm* infection and was identified as an *S* gene for BBC. Characterization of BBC resistant line Acb6 revealed sequence differences in the *GhSWEET10D* promoter, but not in the EBE, that impair activation by Avrb6, possibly relating to resistance mediated by the *b6* gene. Finally, surveys of different *Xcm* field isolates showed that two additional clade III *GhSWEETs* were highly induced during infection. It appears that the evolution of novel TAL effectors that induce these additional clade III *GhSWEET* members may have contributed to the recent re-emergence of BBC in the southern U.S.

The *SWEET* family members are key *S* genes common to several diseases caused by different *Xanthomonas* species[83]. So far, only clade III *SWEET* genes are targeted by pathogens that use TAL effectors to facilitate disease[81, 84]. Members of clade III *SWEET* genes are known to transport sucrose, which likely serves as a carbon source for pathogen. Interestingly, it was recently shown that two clade III *SWEET* genes in *Arabidopsis*, *AtSWEET13* and *AtSWEET14*, can also transport plant hormone gibberellin to modulate growth and development[85]. In rice, *Xoo* can target three clade III *SWEET* genes to cause susceptibility. *Xoo* TAL effector PthXo1 targets *OsSWEET11* [14]; PthXo2 targets *OsSWEET13* [11]; and AvrXa7 targets *OsSWEET14* [13]. In cassava, *Xam* can use $TAL20_{Xam668}$ to target *MeSWEET10a*, another clade III *SWEET* gene[15]. In at least one case, a host *SWEET* gene is targeted by multiple TAL effectors from different strains of the pathogen: in addition to AvrXa7, *OsSWEET14* is targeted by PthXo3, TalC, and Tal5, each from a different *Xoo* strain[12, 13, 86]. The central role of SWEETs in susceptibility is further supported by our data, as we have identified *GhSWEET10* as a clade III *SWEET* gene in cotton targeted by Avrb6. In addition, we observed that several different *Xcm* field isolates strongly induce two additional clade III *GhSWEETs*, *GhSWEET14a* and *GhSWEET14b*, but not *GhSWEET10*, during infections. Since *Xcm*H1005 is a strain derived from an older race, and since these field isolates were collected within the past year, this strongly suggests that new TAL effectors have evolved in these field isolates to target new *SWEETs* and break resistance. The striking differences of the *tal* gene content and the localization of the *pthN2* gene on chromosome or plasmid in *Xcm*H1005 and *Xcm*N1003 are consistent with the notion that

TAL effectors in *Xcm* may evolve rapidly. However, it is also possible that TAL effectors from these recent strains can activate another target, either another *SWEET* gene or a gene in different family, to cause susceptibility. It was revealed that TalC from *Xoo* can still cause disease on rice, despite having its corresponding EBE from *OsSWEET14* artificially mutated, suggesting that preventing TalC-mediated *SWEET14* activation alone is not sufficient to prevent susceptibility[87]. Additionally, even though *GhSWEET10*, *GhSWEET14a*, and *GhSWEET14b* reside in clade III, it is still unclear if they function similarly. Additional biochemical analyses are needed in order to further understand the function of these genes in BBC. Characterizing the spectrum of SWEET-targeting TAL effectors and the relative contributions of their targets to susceptibility, both in the older strains and the recent isolates, is an important future objective.

The central repeat region of TAL effectors determines their DNA binding specificity for targeted gene activation in disease or resistance. Inventory of the repeat sequences is essential to predict important TAL effector targets. However, the repeats are generally impossible to assemble from Illumina sequencing reads and challenge remains even with Sanger sequencing. PacBio sequencing has proven to be an effective approach however[76, 88, 89]. We used PacBio sequencing to reveal the whole repertoire of TAL effectors in *Xcm*H1005 and *Xcm*N1003. We identified 12 TAL effectors encoded by H1005 (six in the chromosome and six in the plasmid) and 9 encoded by N1003 (five in the chromosome and four in the plasmid). In addition to *GhSWEET* genes, another shared class of targets across TAL effectors of different *Xanthomonas* species is transcription factor genes[65, 90, 91]. In at least one case, the CsLOB1 transcription factor

of citrus, and other members of CsLOB1 family also function as *S* genes[65, 92]. It will be interesting to determine whether targets of other TAL effectors in H1005 and N1003 similarly include transcription factor genes, or yet novel types of targets.

In addition to revealing a host factor essential for disease symptom development, our study yields insight into the nature of BBC resistance mediated by *b6*. Initially characterized as recessive[40], genetic linkage mapping analysis suggests that the genetic basis of *b6*-resistance is rather complex and exhibits a quantitative inheritance. In rice, a mechanism of recessive resistance is disruption of TAL effector-dependent activation of an *OsSWEET S* gene due to polymorphism at the EBE[9]. Our data revealed that the induction of *GhSWEET10* by Avrb6 in Acb6 is compromised likely not due to the sequence polymorphism at the EBE, suggesting a novel mechanism underlying *b6*-mediated resistance. To our knowledge, this is the first report of polymorphisms outside the EBE itself that impair activation by a TAL effector. The *b6*-mediated resistance is accompanied by the HR at the site of infection, and this is triggered specifically in the presence of Avrb6 [38, 49]. However, a dTALE activating *GhSWEET10* only results in water-soaking, but not HR in Acb6, suggesting that GhSWEET10 does not mediate Avrb6-induced HR. This is likely that HR is due to the activation of another target by Avrb6 in Acb6, but not in Ac44E. Thus, it seems possible that *b6* comprises at least two genetic determinants, one manifesting in reduced susceptibility (lower expression of the *GhSWEET10 S* gene) and the other mediating the defense associated HR. Perhaps both need to be present to support the full level of resistance.

Nevertheless, our findings suggest a strategy for improve BBC resistance in cotton through genome editing of TAL effector EBEs in *SWEET* genes. Genome editing has been successfully used in rice to mutate the *OsSWEET13 S* gene and generate plants resistant to *Xoo* strains that depend on TAL effector PthXo2 as a major virulence factor[11]. An advantage of this strategy is that relatively minor sequence changes could be made that will destroy the EBE while being unlikely to impinge on native promoter activity and without affecting the coding region, keeping endogenous function of the *SWEET* gene intact[87, 93]. The successful knockout of all homeologs of the *Mlo* gene in hexaploid wheat in a single genome editing experiment indicates that disrupting EBEs in *SWEET* genes in cotton is likely to be feasible[94]. As new TAL effectors important for virulence of this pathogen are characterized in the future, this system could be used to edit the EBEs in the corresponding targets, whether *GhSWEET* genes or other BBC *S* genes yet to be discovered, and to generate plants with broad-spectrum BBC-resistance via pyramiding. Not least, duplicating the Acb6 genotype at *GhSWEET10* in the Ac44E background and in other susceptible genotypes would be a powerful approach to dissect the genetic determinants of *b6*-mediated resistance, potentially by uncoupling loss of susceptibility from the HR. Such characterizations would also likely lead to new strategies for improved control of BBC.

Methods

Plants and bacterial strains and growth conditions

The *Gossypium hirsutum* lines Acala44E (Ac44E), Acalab6 (Acb6), and FM706V, and *G. arboreum* line SA-1415 were grown in 3.5-inch square pots containing Metro Mix 900 soils (Sun Gro Horticulture, Agawam, MA) and tobacco *Nicotiana benthamiana* were grown in 3.5-inch square pots containing Metro Mix 366 soils in a growth chamber at 23°C, 30% humidity and 100 μE $m^{-2}s^{-1}$ light with a 12-hr light/12-hr dark photoperiod. Two-week-old cotton plants were used for *Agrobacterium*-mediated VIGS assay, *Agrobacterium*-mediated transient protein expression assay or protoplast isolation. Plants after inoculation with *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) were transferred into a growth chamber at 28°C, 50% humidity and 100 μE $m^{-2}s^{-1}$ light with a 12-hr light/12-hr dark photoperiod. Bacterial strains used in this study are listed in Supplementary Table 3. *E. coli* and *A. tumefaciens* strains were grown in Luria-Bertani (LB) medium with appropriate antibiotics at 37ºC or 28ºC, respectively. *Xcm* strains were grown in Nutrient Broth (NB) medium with appropriate antibiotics at 28ºC. The *Xcm* field strains were isolated from infected cotton fields in Texas through Koch's postulates with modifications. Water-soaked lesions were excised from infected cotton leaves using a sterile scalpel. The cut lesions were soaked and macerated in 100 μl of sterile water. The bacterial suspension was plated on Nutrient Agar from serial dilutions. Colonies were re-streaked and inoculated into cotton to confirm pathogenicity.

Xcm *inoculations*

The *Xcm* strains were inoculated into 4 ml of NB medium with appropriate antibiotics for overnight. For syringe infiltration, bacterial suspensions adjusted to OD_{600}=0.1 were inoculated into expanded cotyledons of 2-week-old plants or true leaves using a needle-less syringe. Small holes were created with a 25G needle on the underside of the cotyledons to facilitate infiltration. For vacuum infiltration, a final bacterial suspension was made by adjusting the bacterial cultures to OD_{600}=0.01 mixed with 0.04% of Silwet L-77 surfactant solution. Plants that were silenced via VIGS three weeks earlier were submerged into the bacterial suspension inside a desiccator connected to a vacuum pump. The plants were vacuum-infiltrated for 5 mins at 76 mmHg.

DNA isolation for Xcm *whole-genome sequencing*

The *Xcm*H1005 and *Xcm*N1003 DNA used for PacBio sequencing was prepared using a protocol for total genomic DNA isolation with modifications[89]. Bacteria were cultured overnight in 30 ml glucose yeast extract media (2% glucose, 1% yeast extract) in a 250 ml flask at 28ºC on a rotary shaker at 250 rpm, harvested by centrifugation at 2580 x g for 10 min at 4 ºC, then gently resuspended and washed in 20 ml NE buffer (0.15 M NaCl, 50 mM EDTA) twice to remove the extracellular polysaccharide. Cells were then gently resuspended in 2.5 ml solution containing 50 mM Tris, pH 8.0, and 50 mM EDTA, and then 0.5 ml solution containing 25 mM Tris, pH 8.0, 10 µl ReadyLyse (Epicentre) and 50 µl RNase (10 mg/ml). Suspensions were mixed gently by inversion and then incubated on ice for 45 min. Following incubation on ice, 1.0 ml STEP buffer

(0.5% SDS, 50 mM Tris, pH 7.5, 40 mM EDTA, protease K at 2 mg/ml) was added, and the lysate was mixed well by inversion and incubated at 37ºC for 1 h, mixing every 10–15 min. Next, 1.8 ml of 7.5 M ammonium acetate was added and the lysates were mixed gently before being subjected to two extractions with phenol/chloroform (10 ml) and one extraction with chloroform/isoamyl alcohol (24:1, pH 8.0, 10 ml), shaking vigorously by hand to mix, and separating the aqueous and organic phases by centrifugation each time at 10,300 x g for 10 min at 4ºC. Following this, the aqueous phase was transferred to a 14 ml tube and DNA was precipitated by addition of 2 volumes of cold 95% ethanol and gentle, repeated inversion. Once solidified, the DNA was transferred to a 2 ml micro-centrifuge tube using a Pasteur pipette with the tip previously sealed and bent into a hook over a flame. Remaining liquid was then removed by centrifugation at 825 x g for 5 min and the pellet was washed once with 70% ethanol. The remaining liquid was removed as before and the tube left open to dry in a laminar flow hood until the edges of the pellet became glossy in appearance (10–15 min). Finally, the pellet was dissolved in 100 ml TE buffer (10 mM Tris/HCl, 1 mM EDTA, pH 8.0) overnight at 4ºC and then adjusted to a concentration of 1 µg/ml.

PacBio sequencing and genome assembly

DNA library preparation and sequencing was performed according to the manufacturer's instructions by the Genomics Core Facility of the Icahn School of Medicine at Mt. Sinai (New York, NY, USA)[54]. Separately, DNA preparation and electrophoresis to check for small plasmids in either strain that would be lost during

library preparation was using standard methods [50]. Sequencing was conducted to >120 × coverage, which due to improvements in sequencing technology required only two SMRT cells per strain. In all datasets, read-length distribution showed a fat tail, with 20% of coverage after adaptor removal contained in subreads greater than or equal to 15000 bp. The whole genomes of H1005 and N1003 were assembled using HGAP3 (ref. 62) and verified by carrying out local assemblies of reads containing tal genes and then comparing those assemblies to the whole-genome assembly as described in Booher et al[54]. The assemblies conclusively match published Southern blot results using a tal gene probe, as well as restriction enzyme mapping and sequencing or partial sequencing of cosmid clones and subclones [10, 38-40].

The initial assembly of XcmH1005 showed frameshifts in avrB104, avr103, and avrb6 resulting from an extra cytosine (C) in a stretch of six Cs commonly found in the 3' region of tal genes. Since avrb6 is functional, the previously published sequence of avrb6 shows an intact reading frame (ref. 38), and residual errors in homopolymer runs sometimes persist after consensus calling by Quiver and are a known weak point of the error-correction software (https://github.com/PacificBiosciences/GenomicConsensus/blob/master/doc/FAQ.rst), we concluded that the frameshift in avrb6 was such an error, and we corrected it in the final assembly. Likewise, since clones containing avrB104 and avrB103 were functional, conferring avirulence on AcB5 and AcBIn3-containing cotton [40], we concluded that the assembly contained the same error for each of these genes, and corrected them both. No other tal genes in the assembly harbored this or another frameshift within a

homopolymer stretch. Note that the frameshift in pthN', retained in the final assembly, is due to a deletion of a single thymine, not within a homopolymer, and pthN' was shown to be non-functional [10].

Construction of avrb6 *in the expression vectors*

Avrb6 was amplified from *Xcm*H1005 plasmid DNA with primers containing restriction enzymes SpeI at the amino (N) terminus and SmaI at the carboxy (C) terminus (Supplementary Table 4) and ligated into a plant protoplast expression vector *pHBT* with a *CaMV 35S* promoter at the N-terminus and a HA-epitope tag at the C-terminus. To construct the *pCB302* binary vector containing *avrb6* for *Agrobacterium*-mediated transient expression assay in *N. benthamiana*, *avrb6* was released from the *pHBT* vector through digestion with SpeI and SmaI and ligated into the *pCB302* binary vector. All the clones were confirmed by sequencing.

Transient expression assays

Cotton protoplasts were isolated from two-week old cotyledons. Detached cotyledons were cut with a razor blade and digested in an enzyme solution (1.5% cellulose R10, 0.4% macerozyme R10, 0.4 M mannitol, 20 mM KCl, 20 mM MES pH 5.7) supplemented with 2% sucrose for 0.5 h under vacuum. Subsequently, the enzyme solution was incubated without vacuum at room temperature for 12 h. Protoplasts were released by filtering through a 30µm-nylon mesh, washed with W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES pH 5.7) and diluted in MMG solution

(0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES pH 5.7) to a density of 2 x 10^5 cells/ml [95]. 100 µl of protoplasts for each sample were harvested for Western blotting after expressing *avrb6-HA* for 12 h. Avrb6-HA protein was detected by immunoblotting using the α-HA antibody. *Agrobacterium tumefaciens* strain GV3101 containing *pCB302-avrb6-HA* vector or an empty vector was cultured overnight in LB medium containing appropriate antibiotics at 28ºC. Bacteria were harvested by centrifugation and re-suspended with infiltration buffer (10 mM MES, pH 5.7, 10 mM $MgCl_2$, 200 µM acetosyringone) at OD_{600}=0.6. Leaves of four-week-old soil-grown *N. benthamiana* or cotyledons of two-week-old Ac44E cotton were hand-infiltrated using a needleless syringe with *Agrobacterium* cultures. Leaf samples were collected 48 h after infiltration for protein sample and immunoblot analysis.

Cotton RNA isolation for RT-PCR and qRT-PCR analysis

Cotton total RNA was extracted from protoplasts (500 µl protoplasts with a cell density of 2x10^5/ml for each sample) using TRIzol reagent (Invitrogen, USA) or leaves (one cotyledon or one true leaf for each sample) using a Spectrum plant total RNA kit (Sigma-Aldrich, USA) according to the manufacturer's protocol. RNA was then treated with DNase (Invitrogen, USA) to remove genomic DNA. Purified RNA (1 µg) was subsequently reverse transcribed using a first-strand cDNA synthesis kit (Promega, USA) with oligo(dT) as a primer. PCR amplification was performed for 30 cycles of 15 sec at 98ºC, 30 sec at 55ºC, and 45 sec at 72ºC with primers listed in Supplementary Table 4. *GhACTIN* was used as an internal control for RT-PCR. qRT-PCR analysis was

performed using iTaq SYBR green Supermix (Bio-Rad, USA) with a Bio-Rad CFX384 Real-Time PCR System. Expression of each gene was normalized to the expression of *GhUBQ1*.

RNA sequencing and data analysis

A volume of 500 ul of cotton protoplasts ($2x10^5$/ml) were transfected with *35S::avrb6-HA* or an empty vector control. Protoplasts were harvested 12 h after transfection and RNA was extracted by the TRIzol reagent. Three independent biological repeats were performed for the RNA-Seq analysis. For each repeat, equal amount of RNA from two biological replicates was pooled for RNA-Seq library construction. RNA-Seq libraries were prepared using Illumina TruSeqTM RNA Sample Preparation Kit following the manufacturer's instruction, and sequenced on an Illumina HiSeq 2000 platform with 125-nucleotide pair-end reads at Texas AgriLife Genomics and Bioinformatics Service (College Station, TX). RNA-Seq read processing, alignment to the *G. raimondii* genome and differential gene expression analysis were performed as described in Li et al.[96]. Genes with expression fold change ≥ 2 and p-value < 0.05 were considered as significantly different between samples with and without Avrb6.

EBE prediction

All TAL effector EBE predictions were conducted using the TALE-NT 2.0 Target Finder tool[97] to search the promoter regions defined in the *G. raimondii* genome as the 1000 bp upstream of the transcriptional start site plus the 5' UTR, if annotated.

We identified promoter sequences for *G. raimondii* using genome assembly v2.0 and annotation v2.1 [43]. Binding sites passing the standard Target Finder score ratio cutoff of 3.0 were ranked based on both Target Finder output and genomic context using the machine learning classifier presented by Cernadas et al.[69] and updated by Wilkins et al.[98].

Construction of VIGS vector and Agrobacterium *mediated VIGS*

GhSWEET10 was amplified by PCR from cDNA of *G. hirsutum* with primers described in Supplementary Table 4, and inserted into the pYL156 (pTRV-RNA2) vector with restriction enzymes EcoRI and KpnI, resulting in pYL156-*GhSWEET10*. The *Agrobacterium tumefaciens* strain GV3101 carrying the plasmids pTRV-RNA1 and pYL156 (pTRV-RNA2) or pYL156-*GhSWEET10* was harvested and re-suspended with infiltration buffer (10 mM MES, pH 5.7, 10 mM $MgCl_2$, 200 µM acetosyringone) at OD_{600}=0.75. The culture suspension of pTRV-RNA1 was mixed with pYL156 or pYL156-*GhSWEET10* at a 1:1 ratio and hand-infiltrated using needleless syringes into two fully-expanded cotyledons of two-week-old plants [99].

GhSWEET10D *transporter analysis in HEK293T cells*

HEK293T cells were co-transfected with FRET sucrose sensor FLIPsuc90µδ3A and *GhSWEET10D* in 6-well plates and perfusion experiments were performed[84]. After 3.5 min buffer perfusion, cells were perfused with 25 mM sucrose for 4 min followed by 2.5 min buffer perfusion. Images were captured at the following settings: exposure time

800 ms, gain 3, binning 2, time interval 10 s. Fluorescence intensity of selected individual cells was recorded. The quantitative ratio of eCFP and Aphrodite emission at excitation of eCFP was calculated after the background was subtracted. All of ratios were normalized to the initial ratio.

Yeast mutant EBY4000 complementation growth assay

The yeast strain EBY4000 was transformed with the different plasmids following the standard protocol. Spotting assay with tenfold serial dilutions was performed. Yeast cells were diluted to OD_{600}=0.22 in water after cells grew to OD_{600}=0.6, followed by a twofold dilution. Then, tenfold serial dilutions were plated on synthetic medium containing either 2% maltose (as a control), 2% fructose, or 2% glucose. Growth was documented by Gel-doc system after 2-4 days at 30^{o}C.

Transactivation assays

The reporter constructs, *pGhSWEET10D::LUC*, *pGh067700::LUC*, *pGhKBS1::LUC*, *pGhMDR1::LUC*, *pGhHLH1::LUC*, and *pGaSWEET10::LUC*, were constructed by amplifying the promoters of candidate genes (~800 bp upstream of translational start site) from genomic DNA of Ac44E, Acb6, or *G. arboreum* and were fused to a luciferase gene at the C-terminus. Protoplasts were transfected with the effector and reporter construct and expressed at room temperature for 12 h. Protoplasts were then harvested and lysed and the luciferase activity was detected by Glomax Multi-Detection System (Promega, USA) with the luciferase assay substrate (Promega, USA).

UBQ10-GUS was included in all samples as the internal transfection control and the GUS activity was analyzed with a Multilabel Plate Reader (Perkin-Elmer, USA). The *pGhSWEET10D-mEBE::LUC* construct was generated by the site-directed mutagenesis PCR using plasmid DNA of *pGhSWEET10D::LUC* as the template and primers are listed in Supplementary Table 4.

Construction of designer TAL effectors

The dTALEs were constructed using Golden Gate cloning using a complete plasmid kit available through AddGene [100]. Four types of repeats encoding the RVDs NI, NN, NG, and HD that correspond to the respective nucleotide A, G, T, and C were used to assemble the repeat domains of the artificial dTALEs. The dTALEs were transformed into *Xcm* stains via bacterial tri-parental mating by using a helper plasmid, pRK2073 [38]. In addition, the modifier plasmid pUFR054 was used to increase the efficiency of transferring the plasmids from *E. coli* DH5α to *Xcm* strains. The bacterial mating was carried out by mixing the donor, recipient, modifier, and helper strains in a 1:1:1:1 ratio. The mixed culture was plated onto Nutrient Agar medium with appropriate antibiotics. Colonies were confirmed via Sanger sequencing of extracted plasmid DNA.

Measurement of sucrose concentration

Leaves from four-week-old *N. benthamiana* plants were infiltrated with *Agrobacterium* carrying *35S::GhSWEET10D-HA* or an empty vector control at OD_{600}=1.0. Leaves were detached 3 dpi and fully vacuum-infiltrated with sterile

Millipore water. The leaves were blot-dried, cut into small pieces with a razor blade, and placed in a nylon mesh strainer at the top of a 50 ml centrifuge tube. The apoplastic solution was harvested through centrifugation at 3220 x g for 20 mins at 4^{o}C. The concentration of sucrose was determined by using a sucrose assay kit (Sigma-Aldrich).

CHAPTER III

DYNAMIC CO-EVOLUTION OF *XANTHOMONAS*-COTTON INTERACTIONS

Overview

Bacterial blight of cotton caused by *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) is a destructive disease on cotton. While the disease was previously controlled by resistance cultivars, it has surprisingly re-emerged in the U.S. within the last five years. A previous study showed that *GhSWEET10,* a functional plant sucrose transporter, is induced by Avrb6, a TAL effector determining the pathogenicity of *Xcm*H1005, an isolate collected in the 1960s. Additionally, an extensive survey of *GhSWEET* transcriptional responsiveness to different *Xcm* field isolates from Texas revealed that these isolates strongly induced two different *GhSWEETs*, *GhSWEET14a* and *GhSWEET14b*. In this study, we investigated if the *Xcm* field isolates use *GhSWEET14a* and/or *GhSWEET14b* to cause susceptibility. Disease assays combined with qPCR analysis on resistant and susceptible cotton cultivars have suggested a correlation between *GhSWEET14a/b* transcriptional induction and disease severity. Silencing of *GhSWEET14a/b* via virus-induced gene silencing reduced susceptibility to *Xcm*. In order to biologically control bacterial blight of cotton, we employed molecular tools such as the CRISPR-Cas9 system to modify the effector binding element EBE of *GhSWEET10* and a pathogen-inducible microRNA to silence *GhSWEET14a/b* upon infection. These findings advance our understanding of the dynamic disease evolvement in cotton

growing fields and facilitate strategic development for disease resistance improvement in cotton cultivars.

Introduction

Cotton (*Gossypium spp.*) is an important cash crop worldwide that constantly has its yield at risk being constrained due to various biotic factors. Bacterial blight of cotton (BBC), caused by *Xanthomonas citri* subsp. *malvacearum* (*Xcm*), is a destructive disease that has the potential to cause yield losses up to 35% in cotton production. The disease has been well-controlled in the U.S. for nearly a half century. However, it has re-occurred in the southern part of the U.S. for unknown reasons and caused a yield loss of ~$20 million dollars in cotton production in the 2017 growing season as reported by the National Cotton Council of America (http://www.cotton.org).

One major virulence mechanism that is essential for *Xcm* to colonize cotton is to inject transcription activator-like (TAL) effectors which structurally resemble eukaryotic transcription factors [51]. Upon injection into the host via the type III secretion system, TAL effectors localize to the nucleus and directly induce expression of host susceptibility (*S*) genes, or in the case of incompatible interactions, resistance (*R*) genes or "executor" genes [9]. The central DNA binding region consists of 1.5-33.5 copies of near perfect repeats of 33-34 amino acids and is a key domain used to activate transcription of host genes [52, 53]. These repeats are highly conserved with the exception of the 12th and 13th residues of each copy, defined as the repeat variable di-amino acid (RVD). Each RVD targets a specific nucleotide of a continuous sequence in

the host's gene promoter, known as the effector binding element (EBE). The decoding of host DNA recognition by TAL effectors has made it possible to computationally predict TAL effector EBEs in a host genome, design artificial TAL effectors, and identify candidate target genes [8, 101, 102].

Among the *S* genes identified from different hosts, a common target appears to be a gene family of plant sugar transporters known as *SWEET*s. In most cases, directly inducing these *SWEET* genes, particularly those in clade III, appears to be important for causing susceptibility in different pathosystems [64, 103]. To date, there have been at least three *SWEET* genes in rice revealed to be direct targets for TAL effectors in *Xanthomonas oryzae* pv. *oryzae*, the causal agent of bacterial leaf blight of rice. PthXo1 targets *OsSWEET11*, PthXo2 targets *OsSWEET13*, and multiple TAL effectors target *OsSWEET14*. In cassava, *Xanthomonas axonopodis* pv. *manihotis* injects $TAL20_{Xam668}$ to directly activate *MeSWEET10a* to facilitate bacterial blight of cassava disease [104]. These findings demonstrate the evolutionary convergence of different *Xanthomonas spp.* to target *SWEET* gene members to cause susceptibility in different hosts. Studies have suggested that these *Xanthomonas spp.* induce this gene family during infection to cause an over-accumulation of sugars to the apoplast, which provides the bacteria a carbon source possibly for survival and reproduction [26, 64, 103, 105, 106]. Other TAL effector targets characterized to date include the rice bacterial blight *R* genes *Xa27 [66]* and *Xa10* [68], the bacterial leaf streak of rice *S* gene *OsSULTR3;6* [69], the pepper bacterial spot *R* gene *Bs3* [107], and the citrus canker *S* gene *CsLOB1* [65].

Our previous study identified a cotton clade III sugar transporter gene, *GhSWEET10*, is directly activated by Avrb6, a TAL effector from *Xcm*H1005, to cause BBC, the first reported *S* gene in cotton [16]. H1005 is a strain derivative isolated from a cotton field in Oklahoma, U.S. in the 1960s [5]. Previous studies have investigated the avirulence (*avr*) genes in H1005 including Avrb6, which was revealed to significantly promote water-soaking during infection in cotton plants that lack the cognate recessive resistance gene, *b6* [38, 71, 73]. In addition, we previously examined other clade III *GhSWEET*s to investigate their transcriptional change in response to *Xcm* isolates collected within the last 2-3 years in cotton fields [16]. Interestingly, the field isolates were unable to induce *GhSWEET10*, but instead significantly induced two different *GhSWEETs*, *GhSWEET14a* and *GhSWEET14b*. This emphasized the importance of activation of *GhSWEETs* during infection for different *Xcm* strains. Notably, given that TAL effectors appear to be key virulence factors of *Xcm*, this observation suggested that novel TAL effectors had evolved in the field isolates to target different *GhSWEET*s. Thus, manipulating members of this gene family that are key targets for *Xcm* TAL effectors can provide new resistance strategies against BBC.

In this study, we first detected the polypeptides of TAL effectors from different strains of *Xcm* using a TALE-specific antibody and revealed significant changes in the TAL effector assemblies between the older and newer *Xcm* strains. Furthermore, we characterized *GhSWEET14a* and *GhSWEET14b* to determine whether the *Xcm* field isolates use these two *SWEET*s to cause BBC susceptibility. Our data shows that *GhSWEET14a* and *GhSWEET14b* plays a role during infection, as the level of induction

of these two genes was strongly correlated to disease severity and silencing them in cotton via *Agrobacterium*-mediated virus-induced gene silencing (VIGS) reduced susceptibility to *Xcm*. Additionally, modern molecular and genomic technologies were deployed, such as the CRISPR/Cas9 system, to develop strategies to potentially biologically control this disease in cotton fields. These findings advance our understanding of the dynamic disease evolvement in the field and facilitate strategic development for cotton disease resistance improvement against *Xcm*.

Results

Detection of TAL effectors in different Xcm *strains*

*Xcm*H1005, a strain used in a previous study to identify *GhSWEET10* as an *S* gene, was isolated in Oklahoma, U.S. in 1968 [5]. Additionally, this is also known to be a race 1 or 4 derivative [2, 5], unlike the race 18 strain that is commonly identified in infected cotton fields in the U.S. Since TAL effectors appear to be key virulence factors and these two *Xcm* isolates are different races, we hypothesized that their TAL effector assemblies were different. To investigate the diversity of the TAL effector profiling of between different isolates, we performed an immunoblot with *Xcm*H1005 (race 1 or 4), *Xcm*N1003 (possibly race 20), and two likely race 18 strains, MSCT1 (a Mississippi *Xcm* isolate sequenced via Pacific Biosciences (PacBio) by Dr. Shi-En Lu's group) [23], and an *Xcm* isolate collected in 2016 from a cotton field in Texas (TEX) using a TALE antibody. As expected, we observed differences in polypeptide patterns between H1005 and N1003, which was consistent with the PacBio sequencing of these strains (Figure 9).

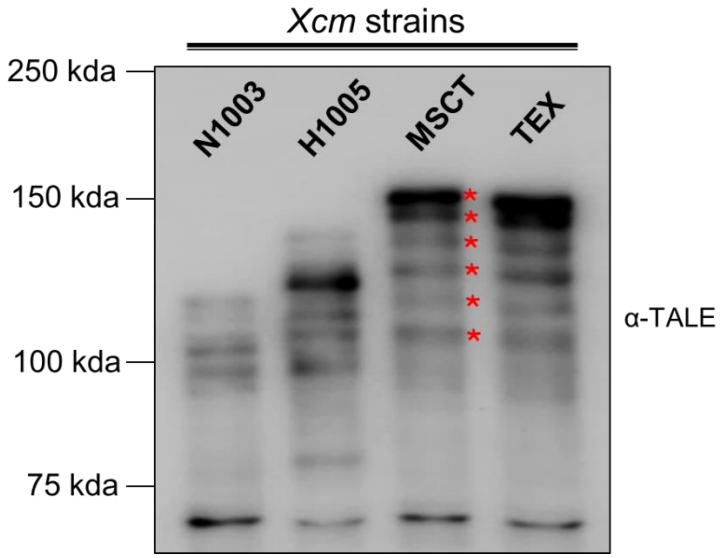

Figure 9. Detection of TAL effectors in different *Xcm* strains.
Xcm strains were cultured overnight in nutrient broth and adjusted to a density of OD_{600} = 0.5 before the addition of SDS buffer. The TAL effectors were detected by immunoblotting with an α-TALE antibody. The red asterisks indicate the annotated TAL effectors from MSCT1 at their predicted size.

Additionally, we observed differences between H1005 and MSCT1. Notably, the sizes of the polypeptides from MSCT1 on the immunoblot matched the predicted size of its annotated TAL effectors [23]. The polypeptide patterns of MSCT1 and TEX appeared to be similar, suggesting that these two isolates possess similar TAL effectors. Interestingly, there appears to be longer TAL effectors in MSCT1 and TEX strains, as polypeptides were detected as high as ~150kDa. The data suggests that there are differences in the TAL effector assemblies between different races.

GhSWEET14a/b *expression correlates to compatible and incompatible interactions*

Our previous study revealed that the current *Xcm* field isolates induce *GhSWEET14a* and *GhSWEET14b* during infection. These isolates are likely race 18 strains. There are commercial cultivars available that are either resistant or susceptible to race 18. We investigated whether the level of induction of *GhSWEET14a/b* was correlated to susceptibility or resistance. To determine this, we inoculated an *Xcm* race 18 isolate into two commercial cotton cultivars shared with us by Bayer CropSciences and measured the transcriptional response of *GhSWEET14a* and *GhSWEET14b*. One was a known susceptible cultivar, Fibermax 2322 (FM2322) and the other was a known resistant cultivar, Fibermax 1830 (FM1830). We observed a higher induction of *GhSWEET14a/b* in the FM2322 cultivar upon infection and a relatively low induction in the FM1830 cultivar (Figure 10a). We also performed a disease assay to measure the bacterial population over a series of days between these two cultivars. There were ~100 fold more bacteria counted from the FM2322 than FM1830 after 5 dpi (Figure 10b).

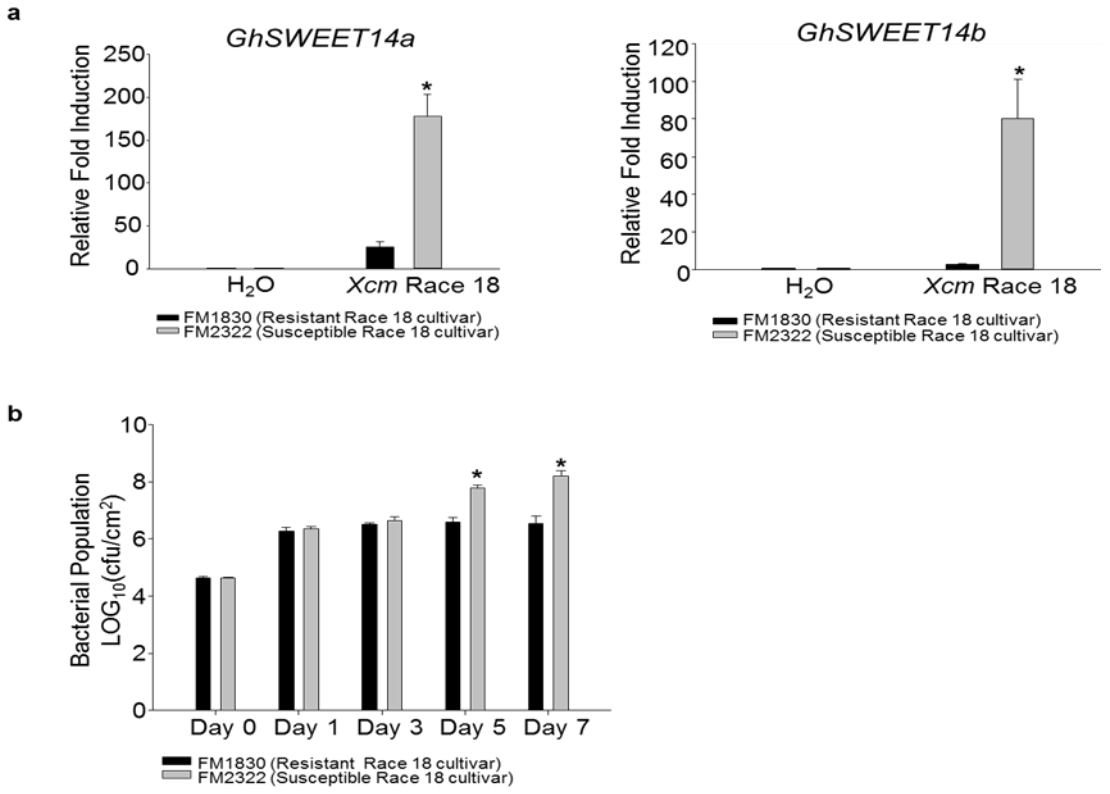

Figure 10. *GhSWEET14a* and *GhSWEET14b* expression correlates to compatible and incompatible interactions in commercial cotton cultivars.

(a) qRT-PCR analysis of *GhSWEET14a* and *GhSWEET14b* in FM1830 and FM2322 plants upon *Xcm* infections. Two-week-old cotyledons were syringe-inoculated with *Xcm* race 18 isolate at OD_{600}=0.1 or sterile water and tissues were collected at 24 hpi. *GhUBQ1* was used as an internal control. The data are shown as mean ± SD (n=3) from three independent repeats. An asterisk indicates significant difference using two-tailed t-test ($p<0.01$) compared to the FM1830 cultivar. (b) *Xcm* bacterial counting on resistant and susceptible commercial cultivars. Two-week old cotyledons were syringe-infiltrated with *Xcm* race 18 isolate at OD_{600} = 0.0001. Samples were harvested at 0, 1, 3, 5, and 7 dpi. The data are shown as mean ± SD from four leaves for each time point. The above experiments were repeated three times with comparable results.

These results suggests that there appears to be a correlation between the level of induction of *GhSWEET14a* and *GhSWEET14b* and the amount of BBC development.

Silencing GhSWEET14a/b *reduces susceptibility to BBC*

To determine if *GhSWEET14a* and *GhSWEET14b* are genetically required for water-soaking during *Xcm* infection in cotton, we silenced these two genes using the *Agrobacterium*-mediated virus-induced gene silencing (VIGS) system before *Xcm* race 18 inoculation. A single VIGS construct, called VIGS-*GhSWEET14ab*, was designed to silence both subgenomes of *GhSWEET14a* (*GhSWEET14aA* and *GhSWEET14a*D) and *GhSWEET14b* because of sequence conservation. Three weeks after VIGS via *Agrobacterium*-infiltration, the *GhSWEET14ab* silenced plants were syringe-inoculated or vacuum infiltrated with *Xcm* race 18. The VIGS-*GhSWEET14ab* plants showed little to no water-soaking upon *Xcm* infection via syringe inoculation compared to the control plants (VIGS-Ctrl) 10 days after inoculation (Figure 11a). Additionally, the VIGS-*GhSWEET14ab* plants consistently showed fewer water-soaked lesions upon infection via vacuum infiltration compared to the VIGS-Ctrl plants (Figure 11b). RT-PCR analysis confirmed reduced induction of *GhSWEET14a* and *GhSWEET14b* in VIGS-*GhSWEET14ab* plants compared to the control (Figure 11c). The data reveals the genetic requirement of *GhSWEET14a* and *GhSWEET14b* for *Xcm* race 18 to cause BBC susceptibility.

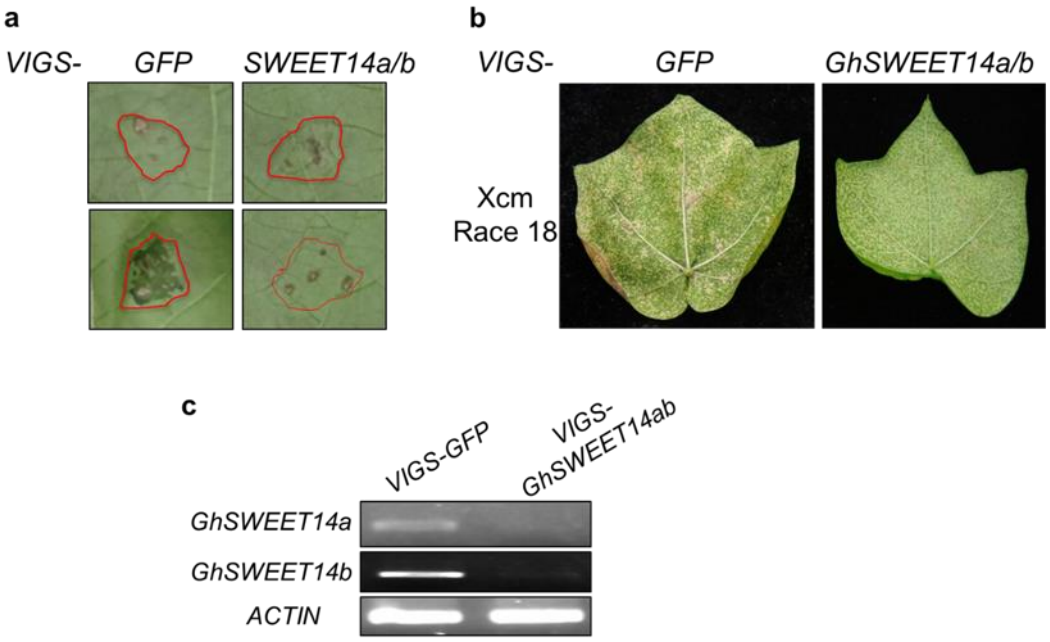

Figure 11. Silencing of *GhSWEET14a/b* in cotton reduces water-soaking caused by *Xcm*.

(a, b) Assessment of water-soaking development caused by *Xcm* on VIGS-ed cotton plants. Cotyledons from two-week-old Ac44E plants were syringe-infiltrated with *Agrobacterium* carrying a VIGS construct to silence both *GhSWEET14a* and *GhSWEET14b* (VIGS-*GhSWEET14a/b*) or a GFP vector control (VIGS-Ctrl). Three weeks later, plants were syringe-infiltrated with *Xcm* race 18 at OD_{600}=0.1 (a) and vacuum-infiltrated with *Xcm*H1005 at OD_{600}=0.001 (b). Images were taken 5 dpi for (a) and two weeks after inoculation for (b). (c) RT-PCR analysis of *GhSWEET14a* and *GhSWEET14b* expression in cotton plants upon VIGS. VIGS assays were done similarly as in (a) and three weeks later, the 2nd true leaf was harvested for RT-PCR analysis before *Xcm* inoculation. *GhACTIN* was used as an internal control. The above experiments were repeated three times with similar results.

Construction of a pathogen-inducible artificial microRNA to control BBC

The data discussed so far suggests that the *Xcm* race 18 isolate as well as other field isolates induces *GhSWEET14a* and *GhSWEET14b* during infection, and that restricting this induction to a relatively low expression subsequently reduces the severity of BBC. Thus, one possible solution to control BBC is to limit the amount of *GhSWEET14a* and *GhSWEET14b* induction during infection. This induction is very likely due to TAL effector(s) binding to a specific region in the promoters of these two *SWEET*s. We decided to take advantage of this knowledge by aiming to fuse the promoters of *GhSWEET14a* and *GhSWEET14b* to an artificial microRNA (amiRNA) that targets both *GhSWEET14a* and *GhSWEET14b* to generate a pathogen-inducible amiRNA. We designed the amiRNA (*amiRNA-SWT14ab*) to target a conserved sequence between both subgenomes of *GhSWEET14a* (*GhSWEET14aA* and *GhSWEET14aD*) and *GhSWEET14b*, then cloned the amiRNA into a plant expression vector by following a previously described protocol [108]. After expressing *amiRNA-SWT14ab* into cotton protoplasts, the protein levels of *GhSWEET14a-HA* and *GhSWEET14b-HA* were analyzed by an immunoblot using an α-HA antibody (Figure 12a). We observed protein reduction of *GhSWEET14a* and *GhSWEET14b* approximately 24 hours after expressing the amiRNA. The results shows that the *amiRNA-SWT14ab* construct can successfully target and silence these two *SWEET* genes. Since the vector is currently under the expression of the CaMV *35S* promoter, we plan to replace it with a synthetic *GhSWEET14a* and *GhSWEET14b* promoter in order to complete the construction of this pathogen-inducible silencing construct (Figure 12b).

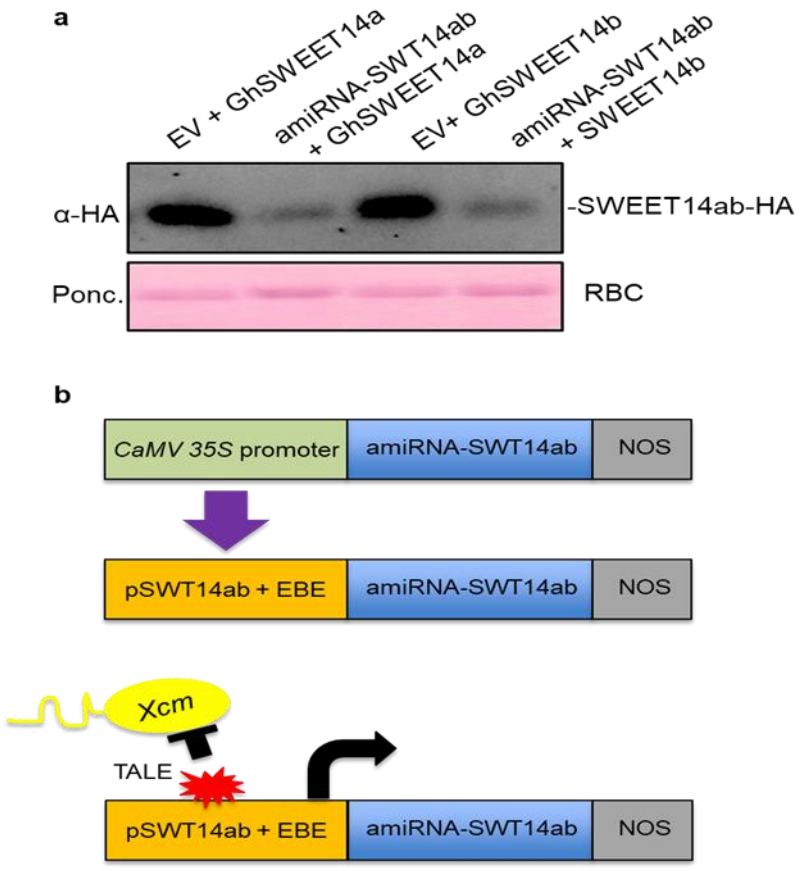

Figure 12. Construction of a pathogen-inducible artificial microRNA to control BBC.

(a) Silencing efficiency of an artificial microRNA targeting GhSWEET14a and GhSWEET14b in cotton protoplasts. Cotton protoplasts were transfected with *GhSWEET14a-HA* or *GhSWEET14b-HA* and *amiRNA-SWT14ab* or an empty vector control (EV). Samples were harvested 24 h after transfection and subjected to immunoblotting with α–HA antibody (top panel). Ponceau S. staining (Ponc.) of total protein served as the protein loading control; RuBisCo (RBC) is shown (bottom panel). (b) Schematic diagram demonstrating a hypothesized method of generating BBC resistant cotton via a pathogen-inducible amiRNA. The current promoter of the *amiRNA-SWT14ab* construct, CaMV *35S*, would be replaced with a synthetic promoter that contains the EBEs of *GhSWEET14a* and *GhSWEET14b* (pSWT14ab). During infection, *Xcm* would use its TAL effector to bind to pSWT14ab and transcriptionally activate the amiRNA, preventing the induction of *GhSWEET14a* and *GhSWEET14b*.

CRISPR/Cas9 targeted mutagenesis of Avrb6 EBE in GhSWEET10D

The CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9 (CRISPR-associated 9) system is a powerful, revolutionary genome editing tool for generating desired mutations in plants [109]. An outcome of this is a targeted loss of function of genes of interest. This is a useful tool in the TAL effector field, as the common forms of resistance against these class of effectors are from their inability to activate *S* genes [11, 14, 93, 110, 111]. *SWEET* genes are common *S* genes of TAL effectors from diverse *Xanthomonas* species [81], and our previous study revealed that *GhSWEET10* is a key *S* gene in BBC since Avrb6 is a key TAL effector in H1005. However since *SWEET* genes are known to be critical for plant and seed development, disrupting their endogenous function should be avoided. We employed the CRISPR/Cas9 system in order to edit the Avrb6 EBE of *GhSWEET10D* and prevent activation by *Xcm* without affecting the gene's function to the plant. The gRNA was designed to target the Avrb6 EBE in the promoter of *GhSWEET10D* (Figure 13a). The gRNA was cloned into a vector that is under the control of two CaMV *35S* promoters, kindly shared by Dr. Yiping Qi, via Golden Gate cloning (*GhSWEET10D-gRNA1*). The *GhSWEET10D-gRNA1* vector and an YFP construct was transformed into cotton callus of cultivar Coker 312 via *Agrobacterium*. Upon successful regeneration of transgenic lines (Figure 13b), the Avrb6 EBE of *GhSWEET10* of these lines was amplified and sequenced to determine the presence of created mutations. We identified two transgenic CRISPR lines, gRNA1-1 and gRNA1-8, that had a 7-bp nucleotide deletion occurred in the Avrb6 EBE (Figure 13c). In order to check that this deletion did not alter the basal

level expression of *GhSWEET10*, we performed a RT-PCR analysis on gRNA1-1 and an YFP transgenic line. We found that there was no difference in *GhSWEET10D* expression between these two lines (Figure 13d). Additionally, we observed no detectable differences in plant phenotype. The results demonstrated the regeneration of cotton transgenic plants with successful genome editing of Avrb6 EBE without altering *GhSWEET10D* endogenous function.

Discussion

BBC is a devastating disease on cotton that has re-emerged within the last few years in the southern U.S. The *Xcm* isolates that are responsible for this re-emergence have not been well-characterized and their host targets to cause disease remain largely unknown. In this study, we identified that these *Xcm* isolates, likely race 18, possess different TAL effectors compared to the relatively older strains, suggesting that these isolates may target different genes instead of *GhSWEET10* to facilitate disease. We also revealed that *GhSWEET14a* and *GhSWEET14b* are induced by these newer *Xcm* isolates and are genetically required for BBC susceptibility. Our findings provided a glimpse of the virulence mechanisms of these *Xcm* isolates and a possible explanation on the re-emergence of BBC. Additionally, we employed modern molecular and genomic technologies to demonstrate new strategies to potentially control BBC in growing fields. We designed an amiRNA that can knock-down the expression of both *GhSWEET14a* and *GhSWEET14b*. Additionally, this amiRNA construct can be modified to have the

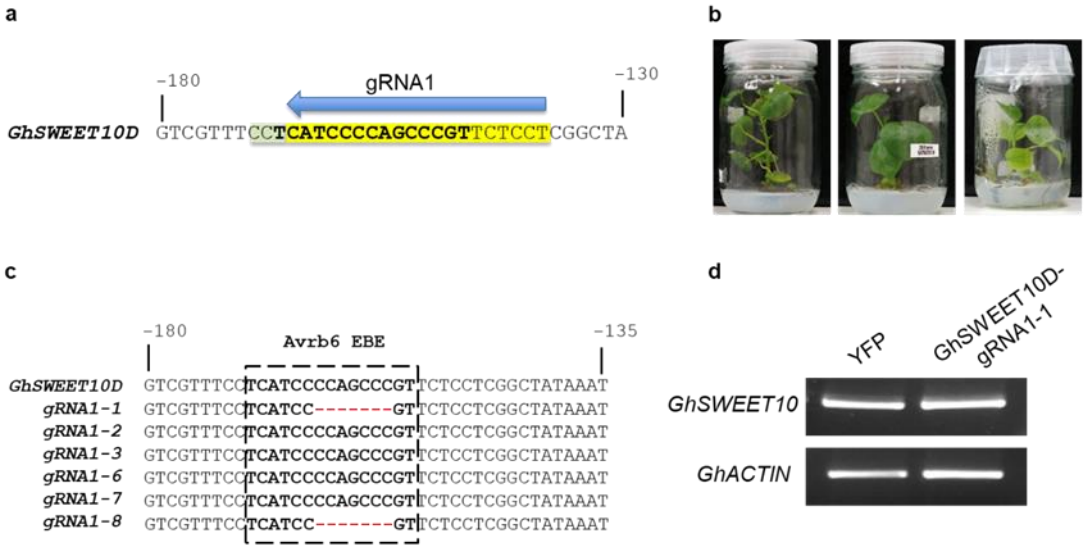

Figure 13. CRISPR/Cas9 mutagenesis of Avrb6 EBE in *GhSWEET10D*.
(a) Targeting region of gRNA in *GhSWEET10D* promoter. The gRNA (gRNA1) was designed to target the sequence highlighted in yellow. The PAM motif was highlighted in green. The Avrb6 EBE is bolded in black. (b) Regeneration of *GhSWEET10D-gRNA1* transgenic plants. (c) Sanger sequencing reveals transgenic lines containing indels of *GhSWEET10D* at the Avrb6 EBE. Nucleotide deletions are shown with red dashes. (d) RT-PCR analysis of *GhSWEET10D* basal level expression in transgenic plants. RT-PCR analysis was performed using two true leaves from *GhSWEET10D-gRNA1-1* or the transgenic YFP plants (control). *GhACTIN* was used as internal control.

construct be driven by a synthetic promoter containing EBEs of *GhSWEET14a* and *GhSWEET14b* in order to generate a pathogen-inducible amiRNA. Furthermore, as proof-of-concept, we were able to regenerate cotton transgenic lines that had the Avrb6 EBE of *GhSWEET10D* edited via the CRISPR/Cas9 system. These molecular strategies employed can empower breeders and farmers with new cotton cultivars that carry stronger resistance against *Xcm* infection.

TAL effectors are key virulence factors of *Xcm* to cause BBC. Avrb6 from H1005 is a major contributor to water-soaking in cotton [73]. The predominant isolate in the U.S. is race 18, which does not appear to have Avrb6 (data not shown). It is still unknown which TAL effector(s) from the race 18 isolates are important to cause water-soaking. A whole-genome assembly of a Mississippi isolate collected in 2011, MSCT1, was recently published and its TAL effector repertoire was revealed [23]. Additional phylogenetic analysis strongly suggested that MSCT1 is a race 18 isolate [22]. We performed an immunoblot analysis of *Xcm* TAL effectors in this study and revealed that MSCT1 and the field isolates from infected cotton fields located in Texas have the very similar polypeptide patterns, suggesting that their TAL effector assemblies are similar. However, despite their protein sizes being similar, it is not yet determined if their TAL effector sequences are identical or if these are even two identical strains. The development of PacBio sequencing provides the capability to assemble full genomes and assemble TAL effectors from different *Xanthomonas* species [89]. Currently, we are performing PacBio sequencing on these *Xcm* isolates found in Texas to reveal their TAL effector assemblies and determine the genetic diversity of *Xcm* isolates across the

country. Furthermore, this will invite additional studies to perform functional characterization of these newly identified effectors. Elucidating the TAL effector assemblies in different *Xcm* isolates could potentially explain the cause of BBC re-emergenence.

Our data indicate that the amount of induction of *GhSWEET14a* and *GhSWEET14b* is strongly correlated to the level of BBC susceptibility and that these two *SWEET*s are genetically required in cotton for *Xcm* to cause water-soaking. It appears that the race 18 isolates are using a similar mechanism as the H1005 strain by inducing a plant sugar transporter in early stages of infection, likely via the injection of TAL effectors. However, the responsible TAL effector(s) involved in directly activating *GhSWEET14a* and/or *GhSWEET14b* have yet to been discovered. Additional experiments are currently on-going to identify which TAL effector(s) from *Xcm* race 18 directly targets any of these two *GhSWEET*s to facilitate disease.

The geographic origins of H1005 and the race 18 strains were relatively close to one another as both strains were isolated in the U.S. (H1005 in Oklahoma, race 18 in Texas). However, the differences in the TAL effector repertoire between H1005 and the current *Xcm* isolates are striking. It has not yet been completely resolved on what caused these effectors to rapidly evolve in a relatively short time. One possibly could be the location of the TAL effectors in the genome. There are 12 TAL effectors in H1005, six on the chromosome and six on plasmid DNA [16], versus 8 TAL effectors in MSCT1, all but one on plasmid DNA [23]. Possessing more TAL effectors on their plasmids could promote more frequent bacterial conjugation between isolates. This may partially

explain why these isolates have different TAL effectors in their genomes despite being isolated in relatively similar locations. Additional whole genome sequencing and assembly of *Xcm* isolates from different areas of the U.S. could potentially reveal the genetic diversity of TAL effectors across a variety of isolates. Another possible scenario that could have caused a shift in the TAL effector assemblies in these isolates is a selection pressure from resistant cotton cultivars. As previously presented, the race 18 and current field isolates did not induce *GhSWEET10* during infection unlike H1005. Since H1005 is a relatively older isolate, perhaps the new resistant cultivars developed either a defense mechanism upon the activation of *GhSWEET10* or have polymorphisms in the Avrb6 EBE of *GhSWEET10*, similar to the recessive alleles in rice such as *xa13* [14, 112]. Future experiments could involve examining the transcriptional response of *GhSWEET10* in resistant cultivars upon *Xcm* infection or sequencing the Avrb6 EBE to determine if new alleles have been developed via breeding.

The availability of draft genome sequences in different cotton species has provided a foundation for cotton researchers to dissect and characterize the function of cotton genes using modern genomic approaches [113]. Despite its complex, allotetraploid genome, the application of the CRISPR/Cas9 system in *G. hirsutum* has been substantially improved, thus providing a powerful tool to study gene function in defense response against pathogens [31-35]. We were able to generate cotton T0 lines with an edited sequence of Avrb6 EBE in *GhSWEET10D* without affecting its basal level expression. Current experiments are on-going to determine if Avrb6-mediated activation of *GhSWEET10D* is disrupted and if the edited cotton lines carries resistance

to Avrb6. Future experiments could also include editing both A- and D-subgenomes as demonstrated in other studies. This proof-of-concept experiment provides an example on using genome editing in cotton to generate novel sources of resistance by making alleles that are insensitive to TAL effector activation. Genome editing the EBEs of additional *GhSWEETs* that are later identified to be *S* genes can provide genetic resources for cotton breeders and farmers for developing new cotton varieties with durable resistance to a variety of *Xcm* isolates.

Methods

Plants and bacterial strains and growth conditions

The *Gossypium hirsutum* (*G. hirsutum*) cultivars Fibermax 2322 (FM2322) and Fibermax 1830 (FM1830) from Bayer CropSciences were grown in 3.5-inch square pots containing Metro Mix 900 soils (Sun Gro Horticulture, Agawam, MA) in a growth chamber at 23^{o}C, 30% humidity and 100 μE $m^{-2}s^{-1}$ light with a 12-hr light/12-hr dark photoperiod. Cotton cultivar Coker 312 (*G. hirsutum*) was germinated on ½ strength MS medium with the growth conditions described later. Two-week-old cotton plants were used for *Agrobacterium*-mediated VIGS assay, *Xcm* disease assay, or protoplast isolation. Plants after inoculation with *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) were transferred into a growth chamber at 28°C, 50% humidity and 100 μE $m^{-2}s^{-1}$ light with a 12-hr light/12-hr dark photoperiod. Bacterial strains used in this study are listed in Supplementary Table 3. *E. coli* and *A. tumefaciens* strains were grown in Luria-Bertani

(LB) medium with appropriate antibiotics at 37ºC or 28ºC, respectively. *Xcm* strains were grown in Nutrient Broth (NB) medium with appropriate antibiotics at 28ºC.

Xcm *inoculations*

The *Xcm* strains were inoculated into 4 ml of NB medium with appropriate antibiotics for overnight. For syringe infiltration, bacterial suspensions adjusted to OD_{600}=0.1 were inoculated into expanded cotyledons of 2-week-old plants or true leaves using a needle-less syringe. Small holes were created with a 25G needle on the underside of the cotyledons to facilitate infiltration. For vacuum infiltration, a final bacterial suspension was made by adjusting the bacterial cultures to OD_{600}=0.001. Plants that were silenced via VIGS three weeks earlier were submerged into the bacterial suspension inside a desiccator connected to a vacuum pump. The plants were vacuum-infiltrated for 5 mins at 76 mmHg.

Cotton RNA isolation for RT-PCR and qRT-PCR analysis

Cotton total RNA was extracted from leaves (one cotyledon or one true leaf for each sample) using a Spectrum plant total RNA kit (Sigma-Aldrich, USA) according to the manufacturer's protocol. RNA was then treated with DNase (Invitrogen, USA) to remove genomic DNA. Purified RNA (1 µg) was subsequently reverse transcribed using a first-strand cDNA synthesis kit (Promega, USA) with oligo(dT) as a primer. PCR amplification was performed for 30 cycles of 15 sec at 98ºC, 30 sec at 55ºC, and 45 sec at 72ºC with primers listed in Supplementary Table 4. *GhACTIN* was used as an internal

control for RT-PCR. qRT-PCR analysis was performed using iTaq SYBR green Supermix (Bio-Rad, USA) with a Bio-Rad CFX384 Real-Time PCR System. Expression of each gene was normalized to the expression of *GhUBQ1*.

Construction of VIGS vector and Agrobacterium *mediated VIGS*

A conserved region of *GhSWEET14aA* and *GhSWEET14aD* and *GhSWEET14b* were amplified by PCR from cDNA of *G. hirsutum* with primers described in Supplementary Table 4, and inserted into one construct of the pYL156 (pTRV-RNA2) vector with restriction enzymes EcoRI and KpnI, resulting in pYL156-*GhSWEET14ab*. The *Agrobacterium tumefaciens* strain GV3101 carrying the plasmids pTRV-RNA1 and pYL156 (pTRV-RNA2) or pYL156-*GhSWEET114ab* was harvested and re-suspended with infiltration buffer (10 mM MES, pH 5.7, 10 mM $MgCl_2$, 200 µM acetosyringone) at OD_{600}=0.75. The culture suspension of pTRV-RNA1 was mixed with pYL156 or pYL156-*GhSWEET14ab* at a 1:1 ratio and hand-infiltrated using needleless syringes into two fully-expanded cotyledons of two-week-old plants [99].

Xcm *bacterial counting*

Xcm strains were collected, washed, and diluted to OD_{600}=0.0001 with sterile water. The bacterial suspensions were inoculated into cotyledons of 2-week-old plants using a needleless syringe. Bacterial counting was performed from four leaves of different plants at each time point. Two leaf discs were ground in 100 µL water and

serial dilutions were plated on nutrient agar medium. Bacterial colony forming units (cfu) were counted two days after incubation at 28^{o}C.

Cotton genomic DNA isolation

Cotton leaves were grinded in liquid nitrogen before adding 600 µL pre-heated CTAB extraction buffer in 1.5 ml tubes. The samples were mixed by vortex and incubated at 60^{o}C for one hour. The samples were subjected to two extractions of chloroform:IAA (24:1), gently inverting the tubes several times to mix, and separating the aqueous and organic phases by centrifugation at 10,000 x g for 10 minutes at 4^{o}C. The aqueous phase was transferred to a new 1.5 ml tube and DNA was precipitated by adding an equal volume of chilled isopropanol. After storing the samples in a -20^{o}C freezer for at least 2 hours, the DNA was pelleted by centrifugation at 10,000 x g for 15 minutes at 4^{o}C, followed by washing once with 70% ethanol. The pellet was dried and dissolved in sterile water.

Immunoblotting for TAL effectors in Xanthomonas

Xcm strains were cultured in 5 ml of nutrient broth for overnight at 28^{o}C. The bacteria were harvested by centrifugation at 1,844 x g for 10 minutes and washed twice with sterile water to remove excess exopolysaccharides. The density of the cultures were adjusted to OD_{600}=0.5 and 120 µL of the bacterial suspension was added to 40 µL 4x SDS buffer (62.5 mM Tris-HCl, pH 6.8, 1% SDS, 0.025% bromophenol blue, 10% glycerol, 1x protease inhibitor for 1 ml of buffer) and boiled at 95^{o}C for 10 min. Proteins

were separated in a 7.5% sodium dodecyl sulfate-poly-acrylamide gel electrophoresis (SDS-PAGE) gel to detect TAL effectors by immunoblotting with TALE anti-serum.

Construction and expression of artificial microRNA

The artificial microRNA (amiRNA) construct for *GhSWEET14ab* was constructed as previously described [108]. Briefly, amiRNA candidates and their cloning primers were obtained from the wmd3 website (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi). The candidate selected was designed to target and silence both *GhSWEET14a* and *GhSWEET14b*. Cloning was performed with pHBT-amiRNA-ICE1 as template as previously described. Constructs of *pHBT-GhSWEET14a-HA* and *pHBT-GhSWEET14b-HA* were constructed by amplifying *GhSWEET14a* and *GhSWEET14b* from cDNA of Ac44E with primers containing restriction enzymes BamHI at the amino (N) terminus and StuI at the carboxy (C) terminus and ligating to the vector. All the clones were confirmed via Sanger sequencing.

For expression, cotton protoplasts were firstly isolated from two-week-old cotyledons as previously described. Protoplasts were transfected with the amiRNA construct and either *HBT-GhSWEET14a-HA*, *HBT-GhSWEET14b-HA*, or an empty vector control and expressed at room temperature for 24 h. Protoplasts were then harvested and GhSWEET14a-HA and GhSWEET14b-HA protein were detected by immunoblotting using the α-HA antibody.

CRISPR vector construction and cotton transformation

The gRNAs were designed by using the CRISPR-P program (http://cbi.hzau.edu.cn/crispr/). The gRNAs were constructed by phosphorylating and annealing the oligos and ligating the annealed oligos to the pTX171 vector that had been linearized with *BsaI* enzyme, resulting in *pTX171-GhSWEET10-gRNA*. The constructed vector was transformed into *Agrobacterium tumefaciens* (LBA4404) and confirmed through sequencing. Cotton transformation was performed in the cotton cultivar Coker 312 by following a previously described protocol with modifications [114, 115]. Hypocotyls were excised from seedlings grown in medium and were infected with *A. tumefaciens* containing the *pTX171-GhSWEET10-gRNA* vector. Explants were co-cultivated on P1-AS medium for three days under light at 25^{o}C and then transferred to selection medium containing hygromycin (P1-c4h15) under light at 28^{o}C in a growth chamber. After 42 days of culture, callus tissue (3 mm or larger) representing an individual transgenic event, growing at cut surface of the explant was excised under a microscope and placed on P7-c4h15 medium and incubated under diffused light at 28^{o}C. After monthly subcultures on MSEm medium somatic embryos were transferred to EG3 medium for germination. Plantlets that emerged were subsequently transferred to MS3 medium for further grown and development before transferring to soil.

CHAPTER IV

CONCLUSIONS

Summary

Cotton is an agro-economically important crop that is constantly threatened by various biotic factors, such as BBC. This disease was previously well-controlled in cotton fields by planting resistant cultivars that were generated from extensive breeding programs. However, this disease mysteriously re-emerged in the U.S. within the last several years and have resulted in substantial yield losses in cotton fields. This became a major concern as the molecular mechanisms of this disease had previously not been well-characterized.

In chapter I, a background of the disease and a literature review that had described the molecular mechanisms of the disease was presented. In addition, I discussed further insights of the disease and how it might have re-emerged in the southern U.S. and possible methods to breed tolerant cotton using current molecular technologies. In chapter II, I used a combination of computational prediction and transcriptomic analysis to reveal a molecular mechanism of BBC and identify the first cotton susceptibility gene targeted by a well-studied strain, *Xcm*H1005. I presented that *Xcm*H1005 can cause disease by using its type III secretion system to secrete a TAL effector, Avrb6, into the plant cell and drive expression of a plant sugar transporter, *GhSWEET10*. In addition, I also showed that *Xcm* field isolates that were recently isolated from infected cotton fields were unable to activate *GhSWEET10*, but instead

activated two different *GhSWEETs*, *GhSWEET14a* and *GhSWEET14b.* This suggested a dramatic evolutionary shift in the TAL effector assembly in the isolates currently found in the field and may be a cause of BBC's re-emergence in the U.S. In chapter III, I further characterized the *GhSWEET* family and its role in susceptibility, and employed strategies to generate tolerant cotton by manipulating *GhSWEETs* that are key susceptibility factors. The findings in chapter III revealed that the *Xcm* field isolates strongly depend on *GhSWEET14a* and *GhSWEET14b* to cause BBC susceptibility. In addition, genome editing was applied to generate cotton with a mutated Avrb6 EBE in the promoter of *GhSWEET10* in order to prevent its activation by *Xcm*H1005 and a pathogen-inducible microRNA construct was assembled which silences *GhSWEET14a* and *GhSWEET14*b upon pathogen infection. In summary, this dissertation advanced the molecular knowledge of BBC by revealing a molecular mechanism of this disease and providing a blueprint to generate BBC tolerant cotton.

In this chapter, I will suggest areas that should be studied in order to further understand the mechanisms of BBC susceptibility and to potentially identify genes that are involved in resistance. I will also discuss additional potential strategies that could be implemented in the future in order to control BBC.

Future studies and Discussion

Sequencing of Xcm *isolates to reveal genetic diversity of TAL effectors*

As described in this manuscript, *Xcm* uses its type III secretion system to inject TAL effectors in plant cell to facilitate BBC susceptibility. Therefore, TAL effectors

appear to be key virulence factors for *Xcm*. Since scientists are able to use computational algorithms to predict the targets of these TAL effectors, revealing their assemblies and repertoires will provide insights on the function of every TAL effector in a strain. However, while there have been genomes of *Xcm* assembled, most of these were sequenced using the Illumina sequencing platform. The Illumina platform, while accurate and powerful, is a technology that sequences genomes using short-reads. Since TAL effectors consist of highly conserved and relatively long, repetitive sequences, using Illumina technology to assemble these effectors would be difficult, if not impossible. Recent advances in PacBio-based genome sequencing have made assembling full genomes and TAL effectors from different *Xanthomonas* species very feasible. At least three *Xcm* genomes have been successfully assembled using PacBio sequencing. Another long-read sequencing technology is a nanopore DNA sequencer called MinION. While the sequencing of this technology is not as accurate as PacBio, the cost of sequencing is relatively lower and the sequencing can be performed simply with a laptop. In addition, Illumina sequencing can be used in combination with MinION to achieve a higher sequencing accuracy. These two powerful sequencing platforms will allow researchers to perform deep-sequencing of the TAL effector repertoire from different *Xcm* strains and characterize their role in BBC.

Based on analysis in chapter III, it's clear that there was a relatively dramatic change in the TAL effector assemblies between *Xcm*H1005 and the current field isolates. Additionally, there appears to be no Avrb6 effector in any of the field isolates. These observations may explain why these isolates induce *GhSWEET14a* and *GhSWEET14b*,

but no longer *GhSWEET10*. Interestingly, these two types of strains are known to be different races. This suggests that the races of *Xcm* differ in pathogenicity likely due to a change in their TAL effector assemblies. It is surprising, however, that there is such a large change in the TAL effector assembly between these two races. It is unknown what factors causes these effectors to change and why relatively rapidly. One hypothesis is since most TAL effectors reside on plasmids (versus chromosomal DNA) that it allows for different bacteria to conjugate sequences frequently. An interesting study to explore this question would be to perform whole-genome sequencing and assembly on every *Xcm* race isolate using a long-read sequencing technology. This would provide insights on how these TAL effectors evolved throughout the years and how rapid this evolution occurred. Additionally, this would reveal the *GhSWEET* genes that each of these races directly targeted to induce susceptibility.

Characterization of the GhSWEET *family*

Plant sugar transporters play significant roles in different areas of plant development where sugar efflux is critical [64, 106]. *SWEET* genes are important in these processes such as phloem loading for long-distance sugar transport, maintenance of sugar homeostasis, and embryo and pollen nutrition and development [81]. *Xanthomonas* species have evolved mechanisms to hijack these sugars in order to grow and reproduce during the infection process in different crop systems. It appears that these *SWEET* genes, particularly those in clade III, are common susceptibility targets for different *Xanthomonas* species [81]. The results in this dissertation also revealed the

same strategy is being used by *Xcm* to infect cotton. However, this is just this family needs to be further characterized in order to enhance our understanding of BBC.

Clade III *SWEETs* are mainly sucrose transporters and are the primary clade out of the four clades that are targeted by *Xanthomonas* species. The phylogenetic analysis of the cotton *SWEET* family in chapter II revealed that there are 18 *GhSWEET* genes in clade III. Three of those were found to be strongly induced during BBC infection in this work, with *GhSWEET10* being directly activated by *Xcm*H1005, and *GhSWEET14a* and *GhSWEET14b* being strongly induced by the field isolates. However, it is still unclear if there are other *GhSWEETs* that play a role during the infection process. It is possible that there could be other *GhSWEETs* that are only slightly induced upon infection or function as secondary targets. An example of this can be observed in the *Pseudomonas syringae-Arabidopsis* pathosystem, as *P. syringae* induces the expression of seven *Arabidopsis SWEETs* but only three of them are dependent on type III effectors [103]. In addition to exploring those *GhSWEET* genes that are induced, those that are down-regulated during infection should also be investigated, where there have been relatively little discussion. Some of these *GhSWEET* genes could be induced and involved in triggering various sugar signaling pathways upon the detection of the pathogen. This would allow the plant to alter its internal environment in order to prevent the pathogen from creating an optimal environment to survive and multiply. During the early stages of BBC infection, *Xcm* could repress these particular *GhSWEET* genes in order to disrupt these sugar signaling pathways. This would allow for the pathogen to create a beneficial environment for survival. Interestingly, studies have shown that when the plant detects a

high accumulation of sugars in the apoplast, sugar transport proteins (STPs) and sucrose uptake transporters (SUTs) are induced, possibly to deprive the excess sugars in the apoplast, resulting in a reduced fitness of the pathogen [116-118]. This is an area that could also be explored in the *Xcm*-cotton pathosystem. Potential studies could focus on performing transcriptomics analysis on genes involved in sugar signaling, both in susceptible and resistant cultivars, or identifying any *GhSWEET*-interacting proteins via biochemical analysis.

Mapping QTLs involved in BBC resistance

There are a number of cotton cultivars that are resistant to specific races of *Xcm* and develop an immediate HR upon infection. This is likely due to the presence of certain *R* genes in the genomes of these cultivars, providing a gene-for-gene interaction with certain *Xcm* strains or races. As discussed in chapter I, although about 20 major *R* genes (or *B* genes) in cotton have been genetically characterized [2], their gene identities and mechanisms remain unknown.

These *B* genes are polygene complexes that were generated via breeding efforts that facilitate BBC resistance in specific, near-isogenic cotton lines[5]. This was generated through multiple backcrossing of the newly identified resistant line to the susceptible Ac44 line [77]. Interestingly, this set of cotton lines has all of the *B* genes integrated, with a single, but different, *B* gene in each of the lines. One of these *B* genes that was discussed in this study is *b6*, a recessive and likely quantitative *r* gene in Acb6 that recognizes Avrb6 and triggers HR [38, 40]. Interestingly, the transcriptional activity

of Avrb6 is required for HR and *b6*-resistance, suggesting that resistance in Acb6 is mediated via activation of target genes by Avrb6 [38]. Since the quantitative and recessive nature of *b6* makes identifying the responsible gene(s) through traditional map-based cloning difficult, and previous studies suggest that its resistance is caused by transcriptional regulation, an approach to identify genes responsible for *b6*-resistance is through transcriptional profiling and computational prediction, a similar approach used to identify *GhSWEET10* as an *S* gene in chapter II. Candidate targets would be identified by those genes differentially expressed in Acb6 and Ac44E in response to Avrb6 and/or genes specifically induced in Acb6 or Ac44E. As observed from the results in chapter II, there are likely multiple genes involved to acquire full *b6*-mediated resistance. Possibly one set of genes could be involved in mediating HR and the other set could be involved in restricting the induction of *GhSWEET10*.

In addition to *b6*, another notable *B* gene is B_{12}. Cotton (*G. hirustum*) lines that carry this gene are resistant to *Xcm* race 18 [39]. Unlike the *b6* gene, B_{12} appears to be a dominant resistance gene. While both the gene identity and resistance mechanisms of B_{12} is unknown, a previous study showed by using mapped restriction fragment length polymorphism markers that it appears to be located on chromosome 14 in the D-subgenome of cultivar S295 [40]. Notably, a later study narrowed down the location of B_{12} to ~900kb using mapped SSR markers CIR246 and JESPR156 [119]. Future studies could involve to continue to fine map this region in cultivar S295 in order to reveal the gene identity of B_{12}. Additionally, these SSR markers could be used to determine if B_{12} is present in other resistant commercial cultivars. Cloning the B_{12} resistance gene would

not only allow researchers to discover a resistance mechanism to BBC, but would also provide breeders and farmers new cultivars that are more tolerant to *Xcm*.

Strategies to biologically control BBC

The results from this dissertation has not only provided some molecular mechanisms of BBC, but also suggested potential strategies to improve cotton resistance to BBC. Examples that were illustrated in this dissertation were using the CRISPR-Cas9 system to edit the Avrb6 EBE of *GhSWEET10* and constructing a pathogen-inducible artificial microRNA which could prevent *GhSWEET14a* and *GhSWEET14b* from being activated by *Xcm* during early stages of infection. The use of molecular tools in breeding efforts would offer the potential to biologically control BBC by creating broad, durable resistance cotton. Given that TAL effectors activate target genes by their EBE region, manipulation of the EBE could potentially provide new strategies to biologically control BBC.

The recent development of the CRISPR-Cas9 system has provided researchers in the field of plant-microbe interactions to edit host genes that are involved in susceptibility. Since genome editing is now starting to become more feasible in cotton, strategies using this technology can be explored to potentially generate BBC-tolerant cotton. Editing the EBEs of targeted genes is a potential approach, since this would only result a relatively small portion of the host promoter being edited, therefore avoiding any disruption with that gene endogenous function. Additionally, there are natural accessions in rice, such as *xa13*, that confer resistance to the cognate TAL effector due to

polymorphisms in the EBE region [14, 111]. Engineering the TAL effector EBEs have been successfully performed in the *Xoo*-rice pathosystem. As additional *GhSWEET S* genes are identified, their TAL effector binding sequences could be edited to generate a new cultivar that is more tolerant, if not resistant to *Xcm*. Additionally, the sequences of the targeted *GhSWEET* genes in different cultivars could be examined to determine if those carrying polymorphisms in the EBE are more resistant to *Xcm*. Potentially, these cultivars could be used to breed new cotton cultivars that produce high quality fiber and carry resistance to BBC.

Another strategy that could be used is to generate a synthetic promoter that can be induced by one or multiple TAL effector(s). In addition to the pathogen-inducible artificial microRNA discussed in chapter III, an alternative could be to pyramid the EBEs of known susceptibility targets in a minimal promoter fused to *R* gene. This would result in the *R* gene being induced upon infection, therefore triggering HR and providing resistance to diverse *Xcm* strains. A similar approach has been used in a study involving citrus canker by creating a construct that has an avirulence gene *avrGf1* driven by a minimal promoter that contains the cognate EBEs targeted by TAL effectors from multiple *Xcc* strains [120]. A limitation to this method, however, is there has not been any BBC *R* genes identified. Alternatively, a known *R* gene from another plant system could be used instead if a BBC *R* gene is not identified. This was successfully demonstrated by Romer et al. by fusing rice promoter *Xa27* to the coding sequence of pepper *R* gene *Bs3* [102]. When *Xoo* injected a TAL effector *avrXa27*, it activated the *Xa27* promoter and triggered *Bs3*-dependent HR in *N. benthamiana*. Nevertheless,

employing a type of artificial pathogen-induced response could be a relatively useful strategy to provide broad spectrum resistance to diverse *Xcm* isolates in the future.

REFERENCES

1. Workman, D., *Cotton exports by country*. Retrieved on August 26, 2018 from http://www.worldstopexports.com/cotton-exports-by-country/
2. Delannoy, E., et al., *Resistance of cotton towards* Xanthomonas campestris *pv.* malvacearum. Annu Rev Phytopathol, 2005. **43**: p. 63-82.
3. *National Cotton Council of America: Pest Management.* Retrieved on August 26, 2018 from http://www.cotton.org/tech/pest/
4. Isakeit, T., *Bacterial blight of cotton.* AgriLife Extension, 2016.
5. Hunter, R.E., L.A. Brinkerhoff, and L.S. Bird, *The development of a set of upland cotton lines for differentiating races of* Xanthomonas malvacearum. Phytopathology, 1968. **58**: p. 830-832.
6. Baldwin, J.C., *Report of the bacterial blight commitee.* Proceedings Beltwide Cotton Production Research Conferences, 1976.
7. Jacques, M.A., et al., *Using ecology, physiology, and genomics to understand host specificity in* Xanthomonas. Annu Rev Phytopathol, 2016. **54**: p. 163-87.
8. Moscou, M.J. and A.J. Bogdanove, *A simple cipher governs DNA recognition by TAL effectors.* Science, 2009. **326**(5959): p. 1501.
9. Hutin, M., et al., *MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility.* Front Plant Sci, 2015. **6**: p. 535.
10. Booher, N.J. and A.J. Bogdanove, *Tools for TAL effector design and target prediction.* Methods, 2014. **69**(2): p. 121-7.

11. Zhou, J., et al., *Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice.* Plant J, 2015. **82**(4): p. 632-643.

12. Streubel, J., et al., *Five phylogenetically close rice* SWEET *genes confer TAL effector-mediated susceptibility to* Xanthomonas oryzae *pv.* oryzae. New Phytol, 2013. **200**(3): p. 808-819.

13. Antony, G., et al., *Rice* xa13 *recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene* Os-11N3. Plant Cell, 2010. **22**(11): p. 3864-3876.

14. Yang, B., et al., Os8N3 *is a host disease-susceptibility gene for bacterial blight of rice.* Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10503-10508.

15. Cohn, M., et al., Xanthomonas axonopodis *virulence is promoted by a transcription activator-like effector-mediated induction of a* SWEET *sugar transporter in cassava.* Mol Plant Microbe Interact, 2014. **27**(11): p. 1186-1198.

16. Cox, K.L., et al., *TAL effector driven induction of a* SWEET *gene confers susceptibility to bacterial blight of cotton.* Nat Commun, 2017. **8**: p. 15588.

17. Read, A.C., et al., *Suppression of Xo1-mediated disease resistance in rice by a truncated, non-DNA-binding TAL effector of* Xanthomonas oryzae. Front Plant Sci, 2016. **7**: p. 1516.

18. Ji, Z., et al., *Interfering TAL effectors of* Xanthomonas oryzae *neutralize* R-*gene-mediated plant disease resistance.* Nat Commun, 2016. **7**: p. 13435.

19. Song, C. and B. Yang, *Mutagenesis of 18 type III effectors reveals virulence function of XopZ(*PXO99*) in* Xanthomonas oryzae *pv.* oryzae. Mol Plant Microbe Interact, 2010. **23**(7): p. 893-902.

20. Furutani, A., et al., *Identification of novel type III secretion effectors in* Xanthomonas oryzae *pv.* oryzae. Mol Plant Microbe Interact, 2009. **22**(1): p. 96-106.

21. Qin, J., et al., *The* Xanthomonas *effector XopK harbours E3 ubiquitin-ligase activity that is required for virulence.* New Phytol, 2018. **220**(1): p. 219-231.

22. Phillips, A.Z., et al., *Genomics-enabled analysis of the emergent disease cotton bacterial blight.* PLoS Genet, 2017. **13**(9): p. e1007003.

23. Showmaker, K.C., et al., *The genome of the cotton bacterial blight pathogen* Xanthomonas citri *pv.* malvacearum *strain* MSCT1. Stand Genomic Sci, 2017. **12**: p. 42.

24. Kemerait, B., et al., *Identification and management of bacterial blight of cotton.* Cotton Incorporated Report, 2017.

25. Xin, X.F., et al., *Bacteria establish an aqueous living space in plants crucial for virulence.* Nature, 2016. **539**(7630): p. 524-529.

26. Aung, K., et al., *The role of water in plant-microbe interactions.* Plant J, 2018. **93**(4): p. 771-780.

27. Yu, X., et al., *From chaos to harmony: Responses and Signaling upon Microbial Pattern Recognition.* Annu Rev Phytopathol, 2017. **55**: p. 109-137.

28. Melotto, M., et al., *Stomatal defense a decade later.* Plant Physiol, 2017. **174**(2): p. 561-571.

29. Panchal, S., et al., *Regulation of stomatal defense by air relative humidity.* Plant Physiol, 2016. **172**(3): p. 2021-2032.

30. Knott, G.J. and J.A. Doudna, *CRISPR-Cas guides the future of genetic engineering.* Science, 2018. **361**(6405): p. 866-869.

31. Zhang, Z., et al., *Simultaneous editing of two copies of* Gh14-3-3d *confers enhanced transgene-clean plant defense against* Verticillium dahliae *in allotetraploid Upland cotton.* Front Plant Sci, 2018. **9**: p. 842.

32. Li, C., et al., *A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (*Gossypium hirsutum L.*).* Sci Rep, 2017. **7**: p. 43902.

33. Janga, M.R., et al., *CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum L.*).* Plant Mol Biol, 2017. **94**(4-5): p. 349-360.

34. Chen, X., et al., *Targeted mutagenesis in cotton (*Gossypium hirsutum L.*) using the CRISPR/Cas9 system.* Sci Rep, 2017. **7**: p. 44304.

35. Wang, P., et al., *High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system.* Plant Biotechnol J, 2018. **16**(1): p. 137-150.

36. Blanvillain-Baufume, S., et al., *Targeted promoter editing for rice resistance to* Xanthomonas oryzae *pv.* oryzae *reveals differential activities for* SWEET14*-inducing TAL effectors.* Plant Biotechnol J, 2016.

37. Kourelis, J. and R.A.L. van der Hoorn, *Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function.* Plant Cell, 2018. **30**(2): p. 285-299.

38. DeFeyter, R., Y.O. Yang, and D.W. Gabriel, *Gene-for-genes interactions between cotton* R *genes and* Xanthomonas campestris *pv.* malvacearum *Avr genes.* Mol Plant Microbe Interact, 1993. **6**(2): p. 225-237.

39. Silva, R.A., et al., *A SSR marker linked to the* B_{12} *gene confers resistance to race 18 of* Xanthomonas axonopodis *pv.* malvacearum *in cotton is also associated with other bacterial blight resistance gene complexes.* Australasian Plant Pathology, 2014. **43**: p. 89-91.

40. Wright, R.J., et al., *D-subgenome bias of* Xcm *resistance genes in tetraploid* Gossypium *(cotton) suggests that polyploid formation has created novel avenues for evolution.* Genetics, 1998. **149**(4): p. 1987-1996.

41. Paterson, A.H., et al., *Repeated polyploidization of* Gossypium *genomes and the evolution of spinnable cotton fibres.* Nature, 2012. **492**(7429): p. 423-427.

42. Wendel, J.F., *New World tetraploid cottons contain Old World cytoplasm.* Proc Natl Acad Sci U S A, 1989. **86**(11): p. 4132-4136.

43. Wang, K., et al., *The draft genome of a diploid cotton* Gossypium raimondii. Nat Genet, 2012. **44**(10): p. 1098-103.

44. Li, F., et al., *Genome sequence of the cultivated cotton* Gossypium arboreum. Nat Genet, 2014. **46**(6): p. 567-572.

45. Li, F., et al., *Genome sequence of cultivated Upland cotton (*Gossypium hirsutum TM-1*) provides insights into genome evolution.* Nat Biotechnol, 2015. **33**(5): p. 524-530.

46. Zhang, T., et al., *Sequencing of allotetraploid cotton (*Gossypium hirsutum L. acc. TM-1*) provides a resource for fiber improvement.* Nat Biotechnol, 2015. **33**(5): p. 531-537.

47. Liu, X., et al., Gossypium barbadense *genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites.* Sci Rep, 2015. **5**: p. 14139.

48. Yuan, D., et al., *The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres.* Sci Rep, 2015. **5**: p. 17662.

49. Chakrabarty, P.K., Y.P. Duan, and D.W. Gabriel, *Cloning and characterization of a member of the* Xanthomonas *avr/pth gene family that evades all commercially utilized cotton* R *genes in the United States.* Phytopathology, 1997. **87**(11): p. 1160-1167.

50. Bogdanove, A.J., S. Schornack, and T. Lahaye, *TAL effectors: finding plant genes for disease and defense.* Curr Opin Plant Biol, 2010. **13**(4): p. 394-401.

51. Boch, J. and U. Bonas, Xanthomonas *AvrBs3 family-type III effectors: discovery and function.* Annu Rev Phytopathol, 2010. **48**: p. 419-436.

52. Mak, A.N., et al., *The crystal structure of TAL effector PthXo1 bound to its DNA target.* Science, 2012. **335**(6069): p. 716-719.

53. Deng, D., et al., *Structural basis for sequence-specific recognition of DNA by TAL effectors.* Science, 2012. **335**(6069): p. 720-723.

54. Schreiber, T. and U. Bonas, *Repeat 1 of TAL effectors affects target specificity for the base at position zero.* Nucleic Acids Res, 2014. **42**(11): p. 7160-9.

55. Doyle, E.L., et al., *TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1.* PLoS One, 2013. **8**(12): p. e82120.

56. Marois, E., G. Van den Ackerveken, and U. Bonas, *The* Xanthomonas *type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host.* Molecular Plant-Microbe Interactions, 2002. **15**(7): p. 637-46.

57. Römer, P., et al., *Plant pathogen recognition mediated by promoter activation of the pepper* Bs3 *resistance gene.* Science, 2007. **318**(5850): p. 645-8.

58. Kay, S., et al., *A bacterial effector acts as a plant transcription factor and induces a cell size regulator.* Science, 2007. **318**(5850): p. 648-51.

59. Chu, Z.H., et al., *Promoter mutations of an essential gene for pollen development result in disease resistance in rice.* Genes & Development, 2006. **20**(10): p. 1250-1255.

60. Yuan, M., et al., *Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive* R *gene* xa13 *in rice.* Plant and Cell Physiology, 2009. **50**(5): p. 947-955.

61. Kay, S., et al., *Detailed analysis of the DNA recognition motifs of the* Xanthomonas *type III effectors AvrBs3 and AvrBs3Deltarep16.* Plant J, 2009. **59**(6): p. 859-871.

62. Huang, S., et al., *The broadly effective recessive resistance gene* xa5 *of rice is a virulence effector-dependent quantitative trait for bacterial blight.* Plant J, 2016. **86**(2): p. 186-194.

63. Romer, P., et al., *Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen,* Xanthomonas oryzae *pv.* oryzae. New Phytol, 2010. **187**(4): p. 1048-1057.

64. Chen, L.Q., SWEET *sugar transporters for phloem transport and pathogen nutrition.* New Phytol, 2014. **201**(4): p. 1150-1155.

65. Hu, Y., et al., Lateral organ boundaries 1 *is a disease susceptibility gene for citrus bacterial canker disease.* Proc Natl Acad Sci U S A, 2014. **111**(4): p. E521-E529.

66. Gu, K., et al., R *gene expression induced by a type-III effector triggers disease resistance in rice.* Nature, 2005. **435**(7045): p. 1122-1125.

67. Wang, C., et al., *XA23 is an executor R protein and confers broad-spectrum disease resistance in rice.* Mol Plant, 2015. **8**(2): p. 290-302.

68. Tian, D., et al., *The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum.* Plant Cell, 2014. **26**(1): p. 497-515.

69. Cernadas, R.A., et al., *Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene.* PLoS Pathog, 2014. **10**(2): p. e1003972.

70. Strauss, T., et al., *RNA-seq pinpoints a* Xanthomonas *TAL-effector activated resistance gene in a large-crop genome.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(47): p. 19480-19485.

71. Yang, Y.N., et al., *Host-specific symptoms and increased release of* Xanthomonas citri *and* Xanthomonas campestris *pv.* malvacearum *from leaves are determined by the 102-bp tandem repeats of PthA and Avrb6, respectively.* Mol Plant Microbe Interact, 1994. **7**(3): p. 345-355.

72. DeFeyter, R., and Gabriel, D.W., *At least six avirulence genes are clustered on a 90-kilobase plasmid in* Xanthomonas campestris *pv.* malvacearum. Mol Plant Microbe Interact, 1991. **4**: p. 423-432.

73. Yang, Y., et al., *Watersoaking function(s) of* Xcm*H1005 are redundantly encoded by members of the* Xanthomonas avr/pth *gene family.* The American Phytopathological Society, 1996. **9**: p. 105-113.

74. Salzberg, S.L., et al., *Genome sequence and rapid evolution of the rice pathogen* Xanthomonas oryzae *pv.* oryzae PXO99A. BMC Genomics, 2008. **9**: p. 204.

75. Darling, A.E., et al., *progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.* PLoS One, 2010. **5**(6): p. e11147.

76. Grau, J., et al., *AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from* Xanthomonas *genomic sequences.* Sci Rep, 2016. **6**: p. 21077.

77. Essenberg, M., et al., *Four near-isogenic lines of cotton with different genes for bacterial blight resistance.* Phytopathology, 2002. **92**(12): p. 1323-1328.

78. Meckler, J.F., et al., *Quantitative analysis of TALE-DNA interactions suggests polarity effects.* Nucleic Acids Res, 2013. **41**(7): p. 4118-4128.

79. Knight, R.L. and J.B. Hutchinson, *The evolution of blackarm resistance in cotton.* J Genet, 1950. **50**(1): p. 36-58.

80. Hou, B.H., et al., *Optical sensors for monitoring dynamic changes of intracellular metabolite levels in mammalian cells.* Nat Protoc, 2011. **6**(11): p. 1818-1833.

81. Eom, J.S., et al., SWEETs, *transporters for intracellular and intercellular sugar translocation.* Curr Opin Plant Biol, 2015. **25**: p. 53-62.

82. Tamura, K., et al., *MEGA6: molecular evolutionary genetics analysis version 6.0.* Mol Biol Evol, 2013. **30**(12): p. 2725-2729.

83. Talbot, N.J., *Cell biology: Raiding the sweet shop.* Nature, 2010. **468**(7323): p. 510-511.

84. Chen, L.Q., et al., *Sucrose efflux mediated by SWEET proteins as a key step for phloem transport.* Science, 2012. **335**(6065): p. 207-211.

85. Kanno, Y., et al., AtSWEET13 *and* AtSWEET14 *regulate gibberellin-mediated physiological processes.* Nat Commun, 2016. **7**: p. 13245.

86. Yu, Y., et al., *Colonization of rice leaf blades by an African strain of* Xanthomonas oryzae *pv.* oryzae *depends on a new TAL effector that induces the* rice nodulin-3 Os11N3 *gene.* Mol Plant Microbe Interact, 2011. **24**(9): p. 1102-1113.

87. Blanvillain-Baufume, S., et al., *Targeted promoter editing for rice resistance to* Xanthomonas oryzae *pv.* oryzae *reveals differential activities for* SWEET14-*inducing TAL effectors.* Plant Biotechnol J, 2017. **15**(3): p. 306-317.

88. Huguet-Tapia, J.C., et al., *Complete genome sequence of the African strain* AXO1947 *of* Xanthomonas oryzae *pv.* oryzae. Genome Announc, 2016. **4**(1).

89. Booher, N.J., et al., *Single molecule real-time sequencing of* Xanthomonas oryzae *genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships.* Microb Genom, 2015. **1**(4).

90. Kay, S., et al., *A bacterial effector acts as a plant transcription factor and induces a cell size regulator.* Science, 2007. **318**(5850): p. 648-651.

91. Sugio, A., et al., *Two type III effector genes of* Xanthomonas oryzae *pv.* oryzae *control the induction of the host genes* OsTFIIAgamma1 *and* OsTFX1 *during bacterial blight of rice.* Proc Natl Acad Sci U S A, 2007. **104**(25): p. 10720-10725.

92. Zhang, J., et al., *Homologs of* CsLOB1 *in citrus function as disease susceptibility genes in citrus canker.* Mol Plant Pathol, 2016.

93. Li, T., et al., *High-efficiency TALEN-based gene editing produces disease-resistant rice.* Nat Biotechnol, 2012. **30**(5): p. 390-392.

94. Wang, Y., et al., *Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew.* Nat Biotechnol, 2014. **32**(9): p. 947-951.

95. Gao, X., et al., *Silencing* GhNDR1 *and* GhMKK2 *compromises cotton resistance to Verticillium wilt.* Plant J, 2011. **66**(2): p. 293-305.

96. Li, F., et al., *Modulation of RNA polymerase II phosphorylation downstream of pathogen perception orchestrates plant immunity.* Cell Host Microbe, 2014.

97. Doyle, E.L., et al., *TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W117-W122.

98. Wilkins, K.E., et al., *TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen* Xanthomonas oryzae *pv.* oryzicola *while strict conservation suggests universal importance of five TAL effectors.* Front Plant Sci, 2015. **6**: p. 536.

99. Gao, X., et al., *Cotton* GhBAK1 *mediates Verticillium wilt resistance and cell death.* J Integr Plant Biol, 2013. **55**(7): p. 586-596.

100. Cermak, T., et al., *Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.* Nucleic Acids Res, 2011. **39**(12): p. e82.

101. Bogdanove, A.J. and D.F. Voytas, *TAL effectors: customizable proteins for DNA targeting.* Science, 2011. **333**(6051): p. 1843-6.

102. Romer, P., et al., *A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20526-31.

103. Chen, L.Q., et al., *Sugar transporters for intercellular exchange and nutrition of pathogens.* Nature, 2010. **468**(7323): p. 527-32.

104. Cohn, M., et al., *Comparison of gene activation by two TAL effectors from* Xanthomonas axonopodis *pv.* manihotis *reveals candidate host susceptibility genes in cassava.* Mol Plant Pathol, 2015.

105. El Kasmi, F., et al., *Microbial effectors and the role of water and sugar in the infection battle ground.* Curr Opin Plant Biol, 2018. **44**: p. 98-107.

106. Julius, B.T., et al., *Sugar transporters in plants: New Insights and Discoveries.* Plant Cell Physiol, 2017. **58**(9): p. 1442-1460.

107. Romer, P., et al., *Plant pathogen recognition mediated by promoter activation of the pepper* Bs3 *resistance gene.* Science, 2007. **318**(5850): p. 645-8.

108. Li, J.F., et al., *Epitope-tagged protein-based artificial miRNA screens for optimized gene silencing in plants.* Nat Protoc, 2014. **9**(4): p. 939-49.

109. Shan, Q., et al., *Targeted genome modification of crop plants using a CRISPR-Cas system.* Nat Biotechnol, 2013. **31**(8): p. 686-8.

110. Liu, Q., et al., *A paralog of the* MtN3/saliva *family recessively confers race-specific resistance to* Xanthomonas oryzae *in rice.* Plant Cell Environ, 2011. **34**(11): p. 1958-69.

111. Chu, Z., et al., *Promoter mutations of an essential gene for pollen development result in disease resistance in rice.* Genes Dev, 2006. **20**(10): p. 1250-5.

112. Yuan, M., et al., *Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive* R *gene* xa13 *in rice.* Plant Cell Physiol, 2009. **50**(5): p. 947-55.

113. Zaidi, S.S., S. Mansoor, and A. Paterson, *The rise of cotton genomics.* Trends Plant Sci, 2018.

114. Rathore, K.S., et al., *Cotton (*Gossypium hirsutum L.*).* Methods Mol Biol, 2015. **1224**: p. 11-23.

115. Rathore, K.S., G. Sunilkumar, and L.M. Campbell, *Cotton (*Gossypium hirsutum L.*).* Methods Mol Biol, 2006. **343**: p. 267-79.

116. Lemonnier, P., et al., *Expression of* Arabidopsis *sugar transport protein* STP13 *differentially affects glucose transport activity and basal resistance to Botrytis cinerea.* Plant Mol Biol, 2014. **85**(4-5): p. 473-84.

117. Yamada, K., et al., *Regulation of sugar transporter activity for antibacterial defense in* Arabidopsis. Science, 2016. **354**(6318): p. 1427-1430.

118. Fotopoulos, V., et al., *The monosaccharide transporter gene,* AtSTP4*, and the cell-wall invertase,* Atbetafruct1*, are induced in* Arabidopsis *during infection with the fungal biotroph* Erysiphe cichoracearum. Plant Physiol, 2003. **132**(2): p. 821-9.

119. Yang, Z., *Fine mapping a bacterial blight resistance gene in the cotton cultivar S295 (Master's Thesis)* Texas Tech University Libraries. 2013.

120. Shantharaj, D., et al., *An engineered promoter driving expression of a microbial avirulence gene confers recognition of TAL effectors and reduces growth of diverse* Xanthomonas *strains in citrus.* Mol Plant Pathol, 2017. **18**(7): p. 976-989.

121. Nature Communications, 2017. **8**.

APPENDIX A

SUPPLEMENTAL FIGURES

Avrb6

5′-MDPIRSRTPSPARELLPGPQPDGVQPTADRGVSPPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFS

DLLRQFDPSLFNTSLFDSLPPFGAHHTEAATGEWDEVQSGLRAADAPPPTMRVAVTAARPPRAKPAPRRR

AAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKY

QDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRN

ALTGAPLNLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPV

LCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLC

QAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA

HGLPPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTLDQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLT

PEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPA

QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDH

LVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMT

QFGMSRHGLVQLFRRVGVTELEARSGTLPPASQRWDRILQASGMKRAKPSPTSTQTPDQASLHAFADSLE

RDLDAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQQSFEVRVPEQRDALHLPLSWRVKRPRTSIGGGLPD

PGTPTAADLAASSTVMREQDEDPFAGAADDFPAFNEEELAWLMELLPQ-3′

Figure A-1. Avrb6 amino acid sequence. The first amino acid of each TAL effector repeat is highlighted in blue. The RVD of each repeat is underlined.

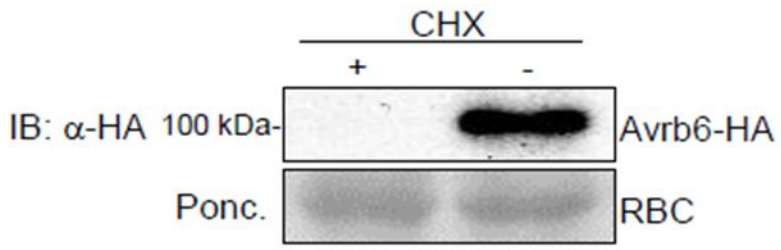

Figure A-2. Cycloheximide (CHX) suppresses Avrb6 protein synthesis in a cotton transient assay. Cotyledons from two-week old Ac44E cotton were infiltrated with *Agrobacterium* carrying *35S::avrb6*-HA at OD_{600} = 0.6 with or without 50µM CHX. Tissues were harvested 4 dpi and subjected to immunoblotting with α-HA antibody (top panel). Ponceau S staining (Ponc) for RuBisCo (RBC) served as the protein loading control (bottom panel).

Figure A-3. Avrb6 EBEs of candidate genes. Avrb6 EBE of each candidate gene is underlined. The sequenced EBE of *Gh067700* bears a deleted nucleotide (indicated with the "C" highlighted in blue) when compared to the predicted Avrb6 EBE.

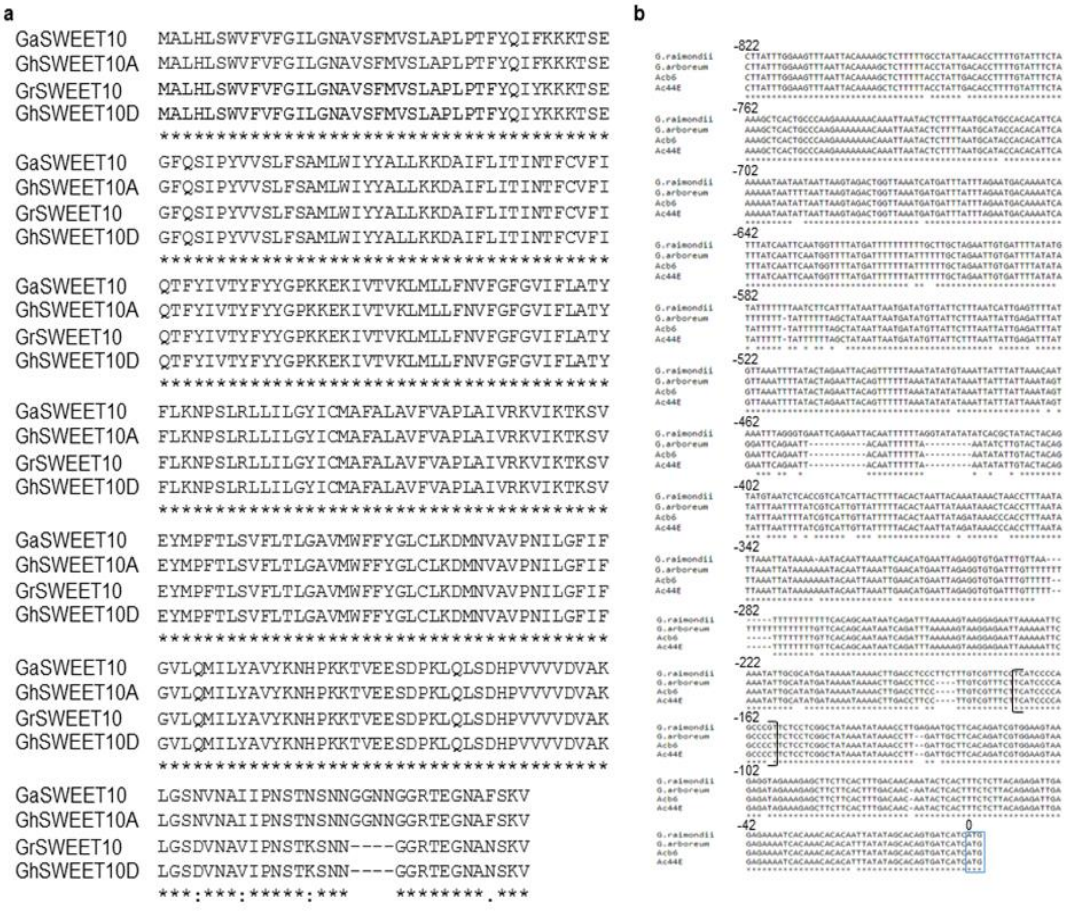

Figure A-4. Sequence alignment of SWEET10 coding sequences and promoters from *G. raimondii, G. arboreum*, and *G. hirsutum*. Alignment was performed with the protein coding sequences (A) and with the promoters using the Clustal Omega website ((http://www.ebi.ac.uk/Tools/msa/clustalo/) (B) with a hierarchical clustering approach. In (B), Avrb6 EBE is highlighted with a black bracket. ATG is highlighted with a blue box. The promoter sequences of *pGhSWEET10A* from A-subgenome in Ac44E and Acb6 are shown here.

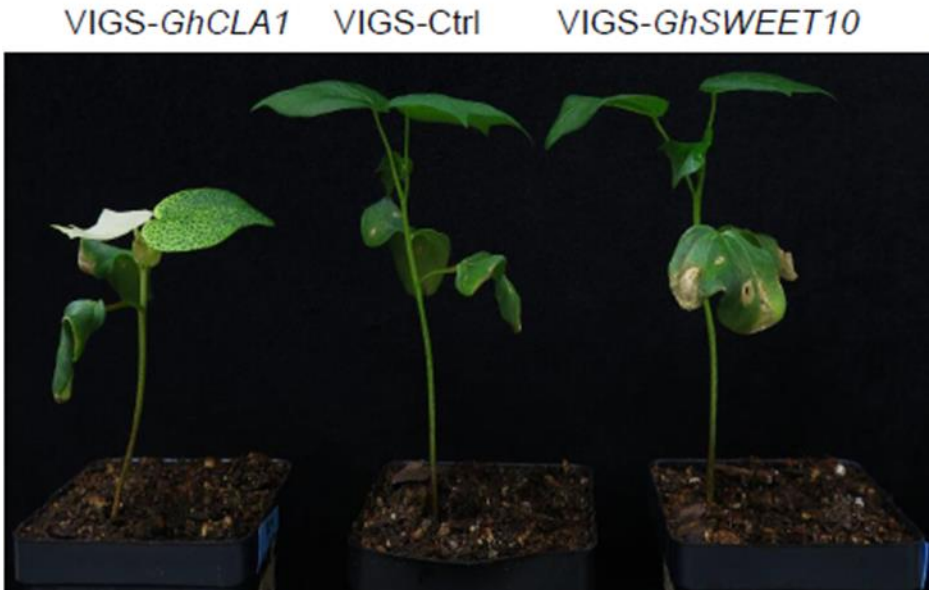

Figure A-5. VIGS-*GhSWEET10* did not affect cotton seedling growth. Pictures were taken at three weeks after inoculation. VIGS-*GhCLA1* is a visual marker for VIGS efficiency.

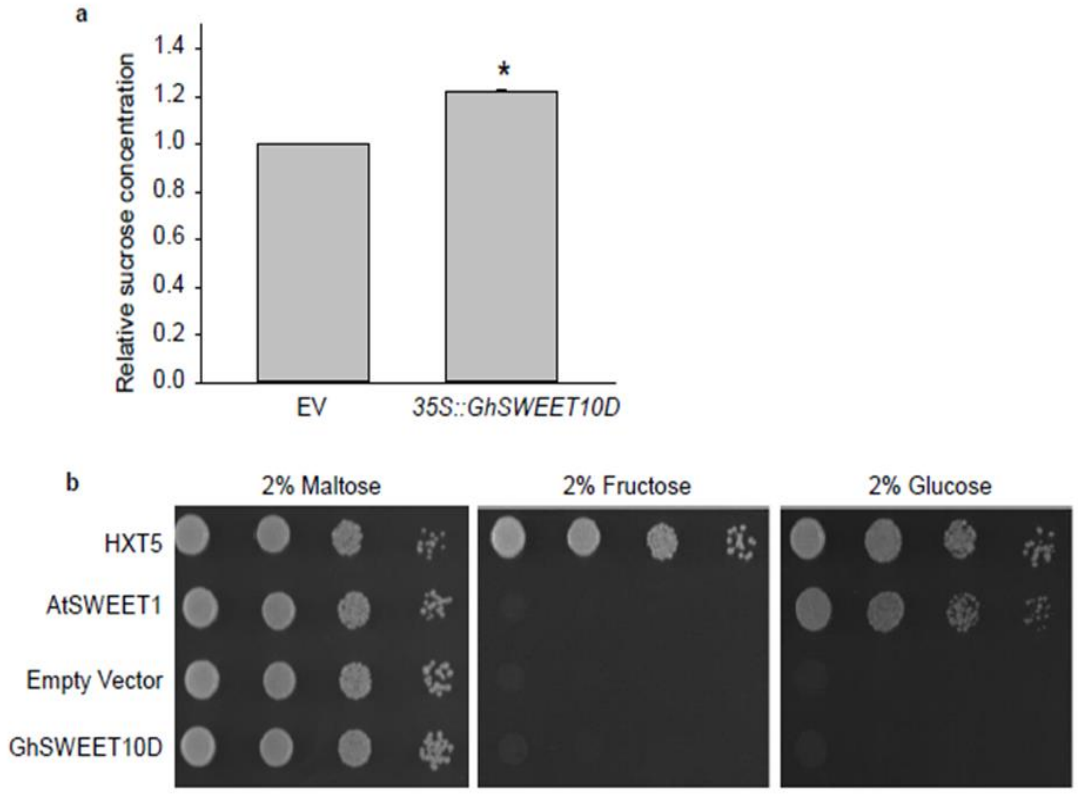

Figure A-6. GhSWEET10D sugar transport activity in *N. benthamiana* and yeast. (A) Expression of *GhSWEET10D* resulted in higher sucrose concentration in apoplasts of *N. benthamiana* leaves. *Agrobacterium* strain carrying *35S::GhSWEET10D* or an empty vector control (EV) was inoculated into *N. benthamiana* leaves. The sucrose content was normalized to a ratio relative to the EV. Error bar is SD and asterisk indicates significant difference using two-tailed t-test ($p<0.05$). The experiment was repeated twice. (B) GhSWEET10D does not complement yeast EBY4000 mutant, a hexose transport-deficient strain. The yeast was grown on the medium supplemented with 2% glucose or 2% fructose as the sole carbon source or 2% maltose as control. HXT5 from yeast and AtSWEET1 from Arabidopsis served as positive controls and empty vector served as a negative control.

Clade I

Gene ID	Gene Name
Gh_A02G0694	GhSWEET1aA
Gh_D02G0740	GhSWEET1aD
Gh_A02G1806	GhSWEET1bA
Gh_D03G1717	GhSWEET1bD
Gh_A11G3285	GhSWEET2aA
Gh_D11G2975	GhSWEET2aD
Gh_D10G0303	GhSWEET2bD
Gh_A11G2655	GhSWEET3A
Gh_D12G2692	GhSWEET3D

Clade II

Gene ID	Gene Name
Gh_A05G3128	GhSWEET4A
Gh_D04G0509	GhSWEET4D
Gh_A05G3127	GhSWEET5aA
Gh_D04G0510	GhSWEET5aD
Gh_A07G0535	GhSWEET5bA
Gh_D07G0604	GhSWEET5bD
Gh_A12G2152	GhSWEET6A
Gh_D12G2328	GhSWEET6D
Gh_A02G0950	GhSWEET7A
Gh_D03G0812	GhSWEET7D
Gh_A13G0907	GhSWEET8aA
Gh_D13G1148	GhSWEET8aD
Gh_A11G2446	GhSWEET8bA
Gh_D11G2763	GhSWEET8bD

Clade III

Gene ID	Gene Name
Gh_A08G0663	GhSWEET9aA
Gh_D08G2730	GhSWEET9aD
Gh_A07G0421	GhSWEET9bA
Gh_D07G0486	GhSWEET9bD
Gh_A12G1747	GhSWEET10A
Gh_D12G1898	GhSWEET10D
Gh_A03G0461	GhSWEET11A
Gh_D03G1078	GhSWEET11D
Gh_A07G0423	GhSWEET12aA
Gh_D07G0487	GhSWEET12aD
Gh_A07G0422	GhSWEET12bA
Gh_A11G0347	GhSWEET13A
Gh_D11G0404	GhSWEET13D
Gh_A04G0861	GhSWEET14aA
Gh_D04G1360	GhSWEET14aD
Gh_D02G1767	GhSWEET14bD
Gh_A01G0160	GhSWEET15A
Gh_D01G0202	GhSWEET15D

Clade IV

Gene ID	Gene Name
Gh_A11G2442	GhSWEET16aA
Gh_D11G2760	GhSWEET16aD
Gh_A13G1540	GhSWEET16bA
Gh_D13G1875	GhSWEET16bD
Gh_A10G1468	GhSWEET17aA
Gh_D10G1709	GhSWEET17aD
Gh_D02G0542	GhSWEET17bD
Gh_D13G1146	GhSWEET17cD
Gh_A13G1434	GhSWEET18A
Gh_D13G1763	GhSWEET18D

Figure A-7. Gene IDs and corresponding names of the cotton GhSWEET family. Tables are separated by different clades of GhSWEET family.

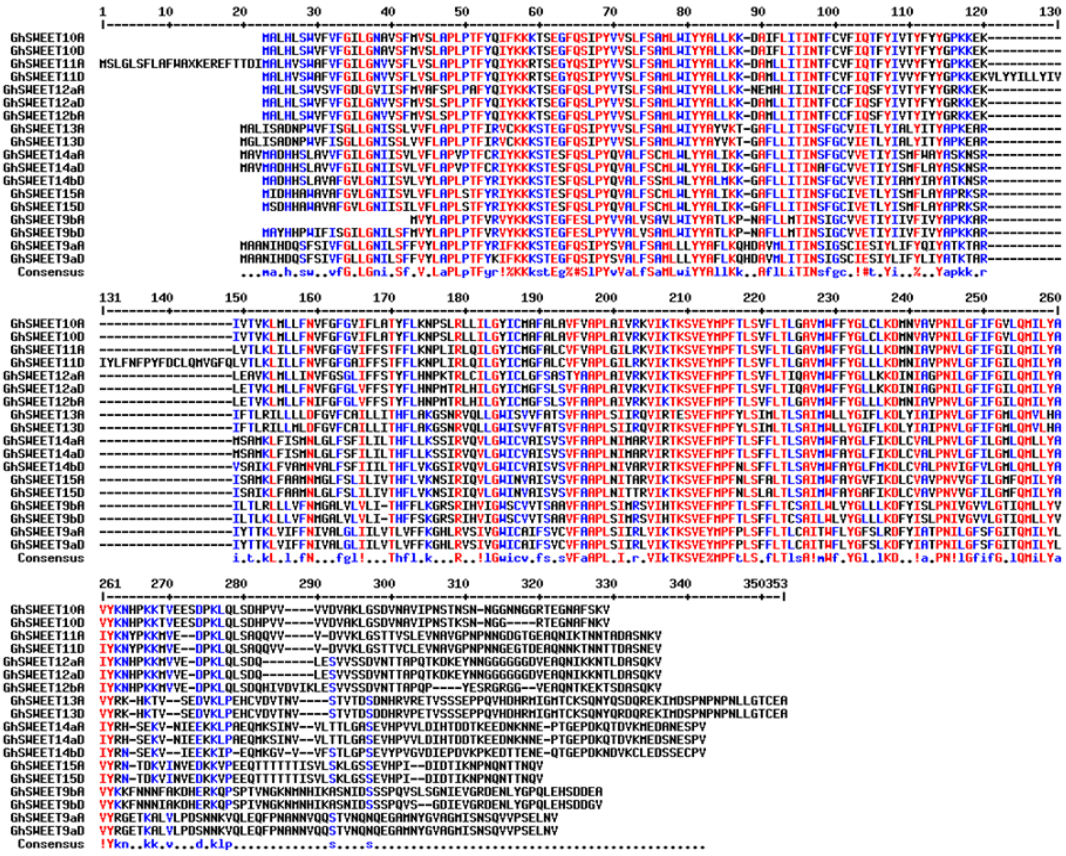

Figure A-8. Amino acid sequence alignment of cotton clade III GhSWEET proteins. Alignment was performed using the Multalin website (http://multalin.toulouse.inra.fr/multalin/) with a hierarchical clustering approach.

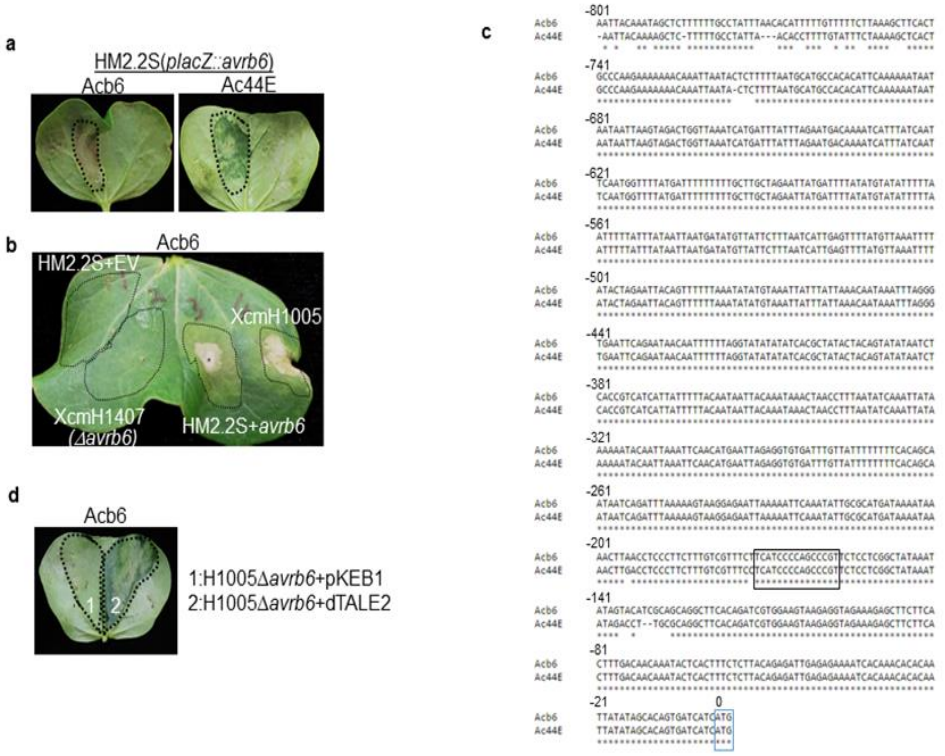

Figure A-9. Differential responses of Avrb6 in Ac44E and Acb6. (A) Avrb6 causes water-soaking on Ac44E, but HR on Acb6. Cotyledons of two-week-old plants were syringe-inoculated with *Xcm*HM2.2S (*placZ::avrb6*) at OD_{600} = 0.1. Pictures were taken at 4 dpi. (B) Strains carrying *avrb6*, XcmH1005 and HM2.2S+*avrb6* cause HR in Acb6. (C) Sequence alignment of the *GhSWEET10D* promoter in Ac44E and Acb6. Alignment was performed using the Clustal Omega website (http://www.ebi.ac.uk/Tools/msa/clustalo/) with a hierarchical clustering approach. Avrb6 EBE is highlighted with a black box. ATG is highlighted with a blue box. (D) dTALE2 causes water-soaking on Acb6. Cotyledons of two-week-old plants were syringe-inoculated with XcmH1005*Δavrb6* carrying dTALE2 or pKEB1 (empty vector) at OD_{600} = 0.1. Pictures were taken at 4 dpi.

APPENDIX B

SUPPLEMENTAL TABLES

Supplementary Table 1. Progressive MAUVE on pXcmH and pXcmN

pXcmH_left end	pXcmH_right end	pXcmN_left end	pXcmN_right end
79681	79764	9909	9992
-3939	-7800	10470	14331
-3463	-3935	14391	14877
16064	25400	16064	25399
25467	25611	25473	25616
59530	63360	35092	38921
7801	16063	0	0
25401	25466	0	0
25612	59529	0	0
63361	79680	0	0
0	0	9993	10469
0	0	14332	14390
0	0	14878	16063
0	0	25400	25472
0	0	25617	35091

Supplementary Table 2. RVD sequences and AnnoTALE names of XcmH1005 and XcmN1003 TAL effectors.

Annotated Name	Gene coordinates	RVD Sequence 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	AnnoTALE Name	Comments on coding sequence
XcmH1005 chromosome				
AvrB5	193705..196791	NI NS HD NI NS NS NS NS HD NI NS HD HD HD	TalDZ1 XcmH1005	
AvrB104	197283..200771	NI HD HD NS HD NG HD HD HD NI NI NI NN NI NG NG N* NG	TalDX1 XcmH1005	
PthN2	228643..231927	NI NG NI NI NS NI NN NI N* NN HD NN NS NG N* NN	TalEG1 XcmH1005	

Annotated Name	**Gene coordinates**	**RVD Sequence** **1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23**	**AnnoTALE Name**	**Comments on coding sequence**
AvrBn	2720814..2723696	NI NG NI HD HD HD HD NG HD HD NG NG	TalEE1 XcmH1005	
AvrB103	2748215..2751499	HD NI NG NI NI NI NG NG NS NG HD HD NI HD N* NN	TalEB2 XcmH1005	
Tal6$_{XcmH1005}$	4979648..4982734	NI NG NI NI NI NG HD HD NG NI HD NI NG NG	TalED1 XcmH1005	
XcmH1005 plasmid (pXcmH)				
AvrB4	8056..11649	NI NG NI NI NS NG NI HD NI NS NS HD HD HD NI NI HD HD NG	TalDY1 XcmH1005	
Avrb6	12430..15516	HD NI NG HD HD NI HD NI NS HD HD HD NN NG	TalEC1 XcmH1005	
AvrB101	32633..36637	HD NG NS NG NI NI NS NG HD NS HD HD NS HD NG NI HD HD NG NS NG HD NG	TalEA1 XcmH1005	
AvrBIn	39817..43617	HD NI NG NI NI NI NG NG NS NG HD HD NI HD NS NG NS NG NG NG NG	TalEB1 XcmH1005	
AvrB102	50873..54367	NI NI NI NG HD NG NI HD NI HD NG NS NG HD NG HD NS NG	TalDW1 XcmH1005	
Avrb7	66226..69822	HD NI NG NI NI NI NS HD HD HD NS HD HD NS NS NG NS NG NG	TalEF1 XcmH1005	
XcmN1003 chromosome				
Tal1'$_{XcmN1003}$	194999..196450	HD	TalCP4 XcmN1003	Only a last, 57 bp (19 aa) repeat; IS element in 3' end
Tal2'$_{XcmN1003}$	2354470..2355501	HD	TalCP5 XcmN1003	Integrase insertion in 2nd repeat
AvrBn	2379939..2382821	NI NG NI HD HD HD HD NG HD HD NG NG	TalEE2 XcmN1003	
Tal4'$_{XcmN1003}$	2830732..2832598	NI HD HD NG HD NG NI NS NS HD	TalCC4 XcmN1003	Frameshift deletion at 11th and 12th repeats
Tal5'$_{XcmN1003}$	4988366..4990750	HD NG NI NI NG HD HD NG NI HD NI NG NG HD NI	TalEI1 XcmN1003	4 bp insertion in last repeat
XcmN1003 plasmid (pXcmN)				
PthN'	26436..29081	NI HD HD NI HD NI NG NI NN HD NI NG N* NN	TalEJ2 XcmN1003	Frameshift in 3' end
Tal7$_{XcmN1003}$	42258..45437	NI HD HD NS NG HD HD N* NN HD NN NS NG N* NN	TalEH1 XcmN1003	

Annotated Name	**Gene coordinates**	**RVD Sequence** **1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23**	**AnnoTALE Name**	**Comments on coding sequence**
PthN	51889..54969	NI HD HD NI HD NI NG NI NN HD NI NG N* NN	TalEJ1 XcmN1003	
PthN2	55295..58444	NI NG NI NI NS NI NN NI N* NN HD NN NS NG N* NN	TalEG2 XcmN1003	Repetitive element insertion in 3' end

Supplementary Table 3. *Xcm* strains and plasmids used in this study.

***Xcm* strains**

Strains	Relevant characteristics	Reference/Source
*Xcm*H1005	Spontaneous Rif^r derivative of *Xcm*H	Yang *et al.* 1994
*Xcm*N1003	Spc^r, Rif^r derivative of *Xcm*N	DeFeyter and Gabriel, 1991
*Xcm*H1407 (H1005Δ*avrb6*)	*avrb6*::Tn5-*gusA*, marker exchange mutant of *Xcm*H1005 (H1005Δ*avrb6*)	Yang *et al.* 1994
H1005 (*placZ::avrb6*)	*Xcm*H1005 with pUFR135	This study
HM2.2S	Mutant strain of H1005 lacking at least six *avr* genes	Chakrabarty *et al.* 1997
HM2.2S (vector)	HM2.2S with pUFR042	This study
HM2.2S (*avrb6*)	HM2.2S with pUFR127	Yang *et al.* 1996
HM2.2S (*placZ::avrb6*)	HM2.2S with pUFR135	This study
HM2.2S+pKEB1	HM2.2S with pKEB1	This study
HM2.2S+dTALE1	HM2.2S with dTALE targeting *GhSWEET10D* in pKEB1	This study
HM2.2S+dTALE2	HM2.2S with dTALE targeting *GhSWEET10* (A&D) in pKEB1	This study
HM2.2S+dTALE3	HM2.2S with dTALE targeting *GhMDR1* in pKEB1	This study
HM2.2S+dTALE4	HM2.2S with dTALE targeting *GhKBS1* in pKEB1	This study
*Xcm*H1005Δ*avrb6*+pKEB1	*Xcm*H1005Δ*avrb6* with pKEB1	This study
*Xcm*H1005Δ*avrb6*+dTALE1	*Xcm*H1005Δ*avrb6* with dTALE targeting *GhSWEET10D* in pKEB1	This study
*Xcm*H1005Δ*avrb6*+dTALE2	*Xcm*H1005Δ*avrb6* with dTALE targeting *GhSWEET10* (A&D) in pKEB1	This study
Xcm1	Field isolate from Plains, TX	This study
Xcm2	Field isolate from Plains, TX	This study
Xcm3	Field isolate from Plains, TX	This study
Xcm4	Field isolate from Plains, TX	This study
Xcm5	Field isolate from Plains, TX	This study

Xcm6	Field isolate from Plains, TX	This study
Xcm7	Field isolate from Matagorda, TX	This study
Xcm8	Field isolate from Lubbock, TX	This study
Xcm9	Field isolate from Seminole, TX	This study
MSCT1	Sequenced *Xcm* field isolate collected in Mississippi 2011	Showmaker *et al.* 2017
TEX	*Xcm* field isolate collected in Texas 2017	This study
Xcm race 18	*race 18 isolate*	This study

Plasmids

Plasmid	Relevant characteristics	Reference/Source
pKEB1	Low copy, gateway-compatible pUFR047 derivative, Gm^r	This study
pRK2073	pRK2013 derivative, npt::Tn7,Kms, Sp^r, Tra^+, helper plasmid	Leong *et al.* 1982
pUFR042	IncW, Km^r, Gm^r, Mob^+, *lacZ*α^+, Par^+	DeFeyter and Gabriel, 1991a
pUFR054	IncP, Tc^r, Mob^+, containing methylases *Xma*I and *Xma*II	DeFeyter and Gabriel, 1991b
pUFR127	5-kb fragment containing *avrb6* in pUFR042	DeFeyter and Gabriel, 1991a
pUFR135	2.9- + -0.5-kb *Bam*HI fragments from *Xcm*H containing intact *avrb6* in pUFR042; *lacZ*::*avrb6*	DeFeyter *et al.* 1993
pTRV-RNA1	pTRV encoding replicase, movement and cysteine-rich protein	Gao *et al.* 2011
pYL156-RNA2	TRV-based VIGS vector, Km^r	Gao *et al.* 2011
pTX171	A single transcript CRISPR/Cas9 vector with 2x CaMV *35S* promoter	From Yiping Qi (unpublished)

Supplementary Table 4. Primers used in this study.

Cloning and mutation primers

Gene	Forward Primer	Reverse Primer
Avrb6	GGACTAGT*ATG*GATCCCATTCGTTCGCG	GTCCCCCGGGCTGAGGCAATAGCTCCATC
PthN	GGACTAGT*ATG*GATCCCATTCGTTCGCG	GTCCCCCGGGCTGAGGCAATAGCTCCCTC
GhSWEET10D	CGGGATCC*ATG*GCTCTTCACTTGTCTTGGG	GAAGGCCTAACTTTGTTGAAAGCATTCCCTTC
pGhSWEET10D	CGGGATCCAAACCACATGGTGGGTGACA	CATGCCATGGCTTCCACGATCTGTGAAGCA

pGh067700	CGGGATCCAGAGGGGCTG ACAGAGGCCCTA	CATGCCATGGTACGATCCG ATCCCCAGTAAAAACAAG
pGhKBS1	CGGGATCCTTCCTTCCAAG AGATGACAAGAGC	CATGCCATGGAGAGATGTG AGCAGAGAGGGA
pGhMDR1	CGGGATCCTCGCTCCCAA AAACAATATAAATCA	CATGCCATGGGGACAGGAC AGTGTTGTGGT
pGhHLH1	CGGGATCCGTCATTCCAC ATATCTCATCCCCA	CATGCCATGGAAGGATGAT TGGATGGTTTGATAGA
pGhSWEET10D -mEBE	CGTTTCCTCATCCCGGGCC CGTTCTCCTC	GAGGAGAACGGGCCCGGG ATGAGGAAACG
VIGS-*GhSWEET10*	GGAATTCACGGCCCAAAG AAAGAAAAGAT	GGGGTACCAGGATGTTTGG AACGGCGAC
pGaSWEET10	CGGGATCCAAACCACATG GTGGGTGACA	CATGCCATGGCTTCCACGA TCTGTGAAGCA
VIGS-*GhSWEET14a*	GGAATTCCGAAAAGCTCG ATTCGCGTCC	GATGCCATGGGGGTGAACT TCAGAGGCACC
VIGS-*GhSWEET14b*	GATGCCATGGGGTTCTGG GTTGGATTTGCG	GGGGTACCTTCTGAGGGGC CTAATGTGC
GhSWEET10D-gRNA1	cggaAGGAGAACGGGCTGG GGATG	aaacCATCCCCAGCCCGTTCT CCT

Note: The restriction enzyme sites are underlined and start codon is italicized. For point mutation primers, the mutation sites are underlined. For gRNA primers, the PAM sequence is in lower-case letters.

RT-PCR Primers

Gene	Forward Primer	Reverse Primer
GhSWEET10	ACGGCCCAAAGAAAGAAAAG AT	AGGATGTTTGGAACG GCGAC
Gh067700	TCACCAACCCTTTGCCTCAAT	AAGGTCTTCATCACCG CCAA
GhKBS1	TCCAGCGAAGATCCCCAGTA	GAACCCTTGACGTTTC GTGC
GhMDR1	CACCCTGGGTACTACATTGTC A	AGTGGCATTATCGCCC ACAA
GhHLH1	ACCAACCAACCAAGAAACAC C	CCCACCTGCTTTGTTG GAGT
GhACTIN	CCTCCGTCTAGACCTTGCTG	TCATTCGGTCAGCAAT ACCA
GaACTIN	CCTCCGTCTAGACCTTGCTG	TCATTCGGTCAGCAAT ACCA

GhSWEET14a	AAAAGCTCGATTCGCGTCC	GGGTGAACTTCAGAGGCACC
GhSWEET14b	GGTTCTGGGTTGGATTTGCG	TTCTGAGGGGCCTAATGTGC

qRT-PCR Primers

Gene	Forward Primer	Reverse Primer
GhSWEET9a	CAGCCAGGATATACACGAC	AAACTGAGAAGATGGCAC
GhSWEET10	GCCATTGTGAGAAAAGTC	TGAAGTACTCCAAAGATG
GhSWEET11	GCATTCTGGGAAACGTTG	ATCCAAAGCATCGCACTG
GhSWEET12a	GCTATCGTGAGAAAGGTC	ACCCTAGTATGTTTGGTC
GhSWEET12b	CCTTGCCAACATTTTATC	AGTAGCATAGCATCCTTC
GhSWEET14a	ACCAGTGCCAACATTCTG	ATCAACGCGTAATACAGC
GhSWEET14b	GGTGTCTTAGGTAACATC	AGCTGAACAATGCCACCTG
GhUBQ1	CTGAATCTTCGCTTTCACGTTATC	GGGATGCAAATCTTCGTGAAAAC